Master Thesis

Discrimination of different crystal types in many-particle-systems using anisotropic scaling indices

Alexander Hartmann

Prof. Gregor Morfill
Dr. Christoph Räth

Masterarbeit

Unterscheidung diverser Kristalltypen in Viel-Teilchen-Systemen anhand anisotroper Skalierungsindizes

Alexander Hartmann

Fakultät für Physik
Ludwig - Maximilians - Universität
München

Betreuer:
Prof. Gregor Morfill
Dr. Christoph Räth

eingereicht am 20. November 2015

Contents

Chapter 1

Introduction

In all experiments and simulations regarding classical many-particle systems, the analysis of the structural properties of the particle distribution is a main issue. In this master thesis, especially the structures forming in "complex" or "dusty" plasmas are investigated. Actually, plasma was considered to be the least orderly state of matter. When small particles – the "dust" – are interspersed into the plasma, however, as it is common in space, e.g. [1] [2], these can arrange in crystalline way, as first described by Hiroyuki Izeki [3] and experimentally observed by Hubertus Thomas et al. [4]. Due to interaction with the plasma, the particles get electrically charged; a sufficiently large charge and resulting electrostatic coupling between the particles induces the forming of crystalline structure. The form and dynamics can easily be manipulated by experimental parameters, such as pressure and particle size. Since these macroscopic crystals are easily observable with high temporal and spatial resolutions, complex plasmas facilitate studies of dynamical processes such as phase transitions and perturbations [5] [6], besides crystals on atomic scales and liquid crystals on macroscopic scales.

Information about local structures in a particle distribution includes the knowledge about order (translational, orientational) or disorder and the shape of the local structures, e.g. the crystal types bcc (body-centered cubic), fcc (face-centered cubic) and hcp (hexagonal close packed), linear structures etc. This knowledge e.g. is important for observing dynamics of the system, such as crystallization fronts [7] and melting and freezing processes [5] [8], or phase transitions between crystal types and energy considerations. There are well-known global methods for analyzing, such as power spectra, pair correlation functions, Minkowski functionals ([9] and references therein) and tensors [10].

In this master thesis, however, I only cover local methods. Such methods e.g. use the local invariants called bond order parameters, which measure the angular distribution of a local structure [11], described later on). Especially, I will address the method using (anisotropic) scaling indices; I will describe its properties and the handling of anisotropy, employ it on artificially created, experimental and simulated data and compare the results to the bond order parameters method.

1.1 About dimensionality and its estimation

The concept of spatial dimensionality does not seem to be a big issue. One-, two- and three-dimensional spatial objects or spaces can easily be imagined and represented, as lines, planes or cuboids, which have a dimensionality $n = 1, 2, 3$, respectively. Also in physics the dimensionality of the phase space where objects "exist in" just refers to the number of its degrees of freedom. For example, a single point can be positioned and move along x-, y- and z-direction; so it has six degrees of freedom and can be described in in a six-dimensional phase space, where there are six independent basis vectors. A larger number of points then can be described in a phase space which has (six $\times$ number of points) dimensions. More general, the term "dimension" can refer to a property needed to describe an object, besides position and momentum, like the dimensions of time, temperature, etcetera. So the meaning of the term "dimension" may extended to the state space, which the object can be fully described in. This state space has the dimensionality of the number of properties needed to describe the object. So the meaning of "dimension" becomes consistent to its meaning besides the field of maths and physics, "dimension" being an aspect of describing a thing.

If a-priori knowledge about the physics of the system is not available (the governing equations), the dimensionality of the system can be obtained empirically by measuring signals from the system, represented each as time series of the form $\{x(t)\}_t$. An embedding of these signals reconstructs the phase space of the system: as proven by Hassler Whitney [12], a d-dimensional manifold can be embedded into a $(2d + 1)$-dimensional Euklidean space, which will be topologically equivalent to the original space. This means when measuring $(2d + 1)$ signals from a system and mapping these into a $(2d + 1)$-dimensional space, the original phase space is reconstructed. As then proven by Floris Takens [13], a delay coordinate map of only one measured signal as well is an embedding. Often, the following form of delay coordinate map is used, τ denotes the delay time:

$$x(t) \to \vec{x}(t) = (x_0(t), x_1(t), x_2(t), \dots x_{2d}(t)), \quad x_p(t) = x(t - p\tau) \tag{1.1}$$

The great advantage of the delay coordinate mapping is that one signal which is very well observable is sufficient for reconstructing the original phase space, other signals which might be hard to measure can be ignored.

For a discrete time series, the delay coordinate τ has to be estimated. One way is calculating the autocorrelation function of the time series $ACF(t', \vec{x}(t))$ and taking the value $t' =: \tau$ of the first zero crossing $ACF(\tau) = 0$. Another way is calculating the mutual information $I(t',\ \vec{x}(t))$ and taking the value $t' =: \tau$ of the first minium of $I(t')$ [14]. The embedding dimension can be estimated by the method of false next neighbors [15]. By a stepwise increasing of the embedding dimension $d \to d{+}1$, the Euklidean distance $R(i,\ j(i))$ between points i and their next neighbors $j(i)$ increases as well $R^d(i,\ j(i)) \to R^{d+1}(i,\ j(i))$. A false next neighbor is identified, if:

$$\frac{R^{d+1}(i,\ j(i))}{R^d(i,\ j(i))} \geq R_{tol}, \quad R_{tol} \geq 10 \tag{1.2}$$

When no more false next neighbors are identified, the embedding dimension is sufficient for reconstructing the phase space.

So the dimensionality of the system is tantamount to the dimensionality of the distribution of the embedded time series $\{\vec{x}(t)\}_t \equiv \{\vec{x}_i\}_i,\ i = 1, \ldots N$ containing N points. But when treating chaotic systems, it turns out that the value of the dimensionality is not longer restricted to integer values and new methods need to be found for estimating the dimensionality (for an introduction to chaos, fractals and their dimensionality also read e.g. [16]).

A "chaotic" system has interesting features: first, the system is deterministic, i.e. the flows of the trajectories are conclusively determined by (non-linear) governing equations. Furthermore, they are aperiodic and extremely sensitive on their initial conditions, i.e. they never repeat exactly or stay on a fixed point, and two different trajectories diverge extremely fast (this has assumed a place in common parlance as "butterfly effect"). Famous examples for chaotic systems are the three-body problem and the Lorenz attractor. The latter one was discovered, when Edward Lorenz described a system governed by three Lorenz equations in the context of modeling convection rolls in the atmosphere [17]. Analyzing the phase space of the Lorenz system, one observes the trajectories of the Lorenz system do not flow onto fixed points or limit cycles, but nevertheless onto a bounded set which is stable against small perturbations, called the Lorenz attractor. An attractor (according to [16, p. 324]) is defined as a closed set with following properties: all trajectories that start within the "basin of attraction" (an open set) flow onto the attractor and remain on it forever; as well all trajectories that start on the attractor remain on it forever (the attractor is "invariant"). Furthermore, the attractor cannot be divided into smaller attractors (the attractor is "minimal"). The Lorenz attractor has additional, chaotic properties, therefore it is called a "strange attractor": the flows on it are aperiodic and extremely sensitive on the initial conditions.

The geometrical structure of the Lorenz attractor is very complex. It is not possible to assign a number of values to a point on the attractor in order to set its coordinates on the attractor. E.g. on a line, an one-dimensional object, one value is sufficient to localize a point, i.e. the point will get assigned a coordinate of e.g. $x = 5.7$. On a plane, two coordinates are enough; etc. Since this is not possible for the Lorenz attractor, the concept of dimensionality changes in this context. It turned out that the dimensionality of the Lorenz attractor lies between 2 and 3. This non-integer dimensionality is called a "fractal dimensionality", the attractor a "fractal". A fractal in general is characterized by a very complex geometrical structure with fine structures at all scales, often with self-similarity (also see the von Koch curve example in [16, p. 404ff]).

For estimating the dimensionality of a fractal, there are different methods. These apply following idea: When performing a "measurement at scale ϵ" [16, p. 409] ("measurement" still needs to be defined), how does these vary for $\epsilon \to 0$? According to this, a power law is assumed, where the dimensionality of the fractal equals the exponent in the power law.

One method is the "box-counting dimensionality", e.g. see [18, p. 41ff]. Here the point

distribution is covered with a grid of grating width ϵ. The number of cells $N(\epsilon)$ is counted which is needed for total covering. By assuming the power law

$$N(\epsilon) \sim (\frac{1}{\epsilon})^{D_{box-counting}} \tag{1.3}$$

the box-counting dimension $D_{box-counting}$ then is determined by the limit $\epsilon \to 0$:

$$D_{box-counting} = -\lim_{\epsilon \to 0} \frac{\ln N(\epsilon)}{\ln \epsilon} \tag{1.4}$$

Another method uses the scaling behaviors of the points of the distribution, the "correlation dimension" [19]. In a dense distribution of integer dimensionality, $n = 1, 2, \ldots$, the number of points N contained within a sphere of radius r around some point i just scales with the dimension n: so in a two-dimensional space, the scaling will have the power law $N \sim r^2$, in a three-dimensional space $N \sim r^3$. This method can be generalized to fractal dimensions. Compared with the box-counting dimensionality, here only the positions $\{\vec{x}_i\},\ i = 1, \ldots N$ of the points in the fractal are needed, and the computing time is far shorter. For the correlation sum

$$c(r) = \lim_{N \to \text{inf}} \frac{1}{N^2} \sum_{i,j=1}^{N} \Theta(r - \|\vec{x}_i - \vec{x}_j\|) \tag{1.5}$$

a power law of the form $c(r) \sim r^{D_{corr}}$ is assumed. Here Θ denotes the Heaviside-function: $\Theta(x - a) = 1$ for $x \geq a$ and $\Theta(x - a) = 0$ for $x < a$. r denotes a radius. The power law then yields the correlation dimension D_{corr}:

$$D_{corr} = \lim_{r \to 0} \frac{\mathrm{d} \ln c(r)}{\mathrm{d} \ln r} \tag{1.6}$$

A modification of the correlation dimension, which is a global measure, is the "pointwise dimension" which measures the local scaling behavior for a certain point i of the distribution $\{\vec{x}_i\},\ i = 1, \ldots N$. The formulas are analogue, but note that the sum over index i is omitted in the correlation sum:

$$c_i(r) = \lim_{N \to \text{inf}} \frac{1}{N} \sum_{j=1}^{N} \Theta(r - \|\vec{x}_i - \vec{x}_j\|) \tag{1.7}$$

As well a power law of the following form is assumed:

$$c_i(r) \sim r^{D_{pointwise}} \tag{1.8}$$

There are methods for the backwards conclusion on the global dimensionality by suitable averaging. In this thesis, however, I restrict on local structures, and used an algorithm based on an extension of the pointwise dimensionality, namely scaling indices.

1.2 Introduction to scaling indices

In the following I will describe the concept of (anisotropic) scaling indices, which is an extension of the pointwise dimensionality. This introduction to scaling indices is guided by reference [20].

In a n-dimensional point distribution containing N points $\{\vec{x}_i\}$ with $i = 1 \dots N$ the local scaling index $\alpha_i(r)$ of point i gives information about the scaling behavior of the neighborhood of point i. In the following index i means an arbitrary but fixed point of the distribution. Scaling behavior means how the number of neighboring points $N_{neighbors,i}$ surrounding point i scales with a radius-like parameter r:

$$N_{neighbors,i}(r) \sim r^{\alpha_i(r)} \tag{1.9}$$

So α_i be seen as a local dimensionality, which may as well be non-integer (fractal dimension). In a dense and homogeneous n-dimensional point distribution, $\alpha_i(r)$ will attain the value n for small r:

$$\lim_{r \to 0} \alpha_i(r) = n \quad \forall i \tag{1.10}$$

This is tantamount to the case of a homogeneous and infinite n-dimensional point distribution which is not dense, where $\alpha_i(r)$ will attain the value n for large r:

$$\lim_{r \to \inf} \alpha_i(r) = n \quad \forall i \tag{1.11}$$

The second case can be observed in the results for the point distributions analyzed in this thesis, with the constraint that those ones were not infinite but border effects appeared. Nevertheless, $\alpha_i(r)$ attains the value $n = 2$ or $n = 3$ according to two- or three-dimensional point distributions, respectively, for r equaling a few average point distances, as you will see later.

The number of points $N_{neighbors,i}$ can be also expressed as point density by the cumulative distribution function $\rho_i(r)$:

$$\rho_i(r) = \sum_{j=1}^{N} s_r(d_{ij}) \tag{1.12}$$

$\rho_i(r)$ is dependent on the shaping function $s_r(d_{ij})$, which describes the way of weighting the distance d_{ij} between points i and j. Most easily, the shaping function can be chosen as a Θ-function. The Θ-function, also called Heaviside-function, $\Theta(x - a)$ equals 1 for all $x \geq a$ and equals 0 for all $x < a$. So when setting up the shaping function as follows:

$$s_r(d_{ij}) = \frac{1}{N} \sum_{j=1}^{N} \Theta(r - d_{ij}) \tag{1.13}$$

then all points in the vicinity of point i within a sphere of radius r are counted (fig 1.1).

The distance between points i and j is denoted by $d_{ij} = \|\vec{x}_i - \vec{x}_j\|$. The 2-norm of $\|\vec{x}\|$ can be expressed using the inner product in its general form $\vec{x}^T \cdot A \cdot \vec{x}$, where A is some metric tensor (Hermitian $A = \bar{A}^T$ and positive definite):

$$\|\vec{x}\|_A = \left(\vec{x}^T \cdot A \cdot \vec{x}\right)^{\frac{1}{2}} \tag{1.14}$$

First, A is set to the identity $A = 1$, which makes the general form of the 2-norm turn into the standard Euklidean norm, as follows, with p denoting the single dimensions $p = 1, \ldots n$ and $\vec{x}_{ij} = \vec{x}_i - \vec{x}_j$:

$$d_{ij} = \left(\vec{x}_{ij}^T \cdot \vec{x}_{ij}\right)^{\frac{1}{2}} = \left(\sum_{p=1}^{n} (x_{i,p} - x_{j,p})^2\right)^{\frac{1}{2}} \tag{1.15}$$

Furthermore, I assume a power law for $\rho_i(r)$ as follows (eq 1.16).

$$\rho_i(r) \sim r^{\alpha_i(r)} \tag{1.16}$$

According to this power law, $\alpha_i(r)$ is calculated as the partial derivate of $\ln \rho_i(r)$ with respect to $\ln r$:

$$\alpha_i(r) = \frac{\partial \ln \rho_i(r)}{\partial \ln r} \tag{1.17}$$

An obvious disadvantage of using a Θ-function as shaping function of $\rho_i(r)$ is the resulting stepwise shape of $\rho_i(r)$ (fig 1.1), which makes $\rho_i(r)$ non-differentiable and requires a scaling range $[r_1, r_2]$ for computing α_i:

$$\alpha_i(r_1, r_2) = \frac{\ln \rho_i(r_2) - \ln \rho_i(r_1)}{\ln r_2 - \ln r_1} \tag{1.18}$$

1.3 Weighted scaling indices

Weighted scaling indices redress the problem of the missing differentiability of $\rho_i(r)$ by modification of the shaping function $s_r(d_{ij})$. Using an exponential function instead of a Θ-function is a convenient way of doing this:

$$\rho_i(r) = \sum_{j=1}^{N} e^{-(\frac{d_{ij}}{r})^q} \tag{1.19}$$

Now all points in the set $\{\vec{x}_j\}$, $j = 1 \ldots N$ have impact on the calculation. $\rho_i(r)$ allocates smoothly a value between 0 and 1 to all points in the set and sums them up. The shape of the exponential function is controlled by the parameters r and q. r controls the size of the exponential function: at $d_{ij} = r$ the point j gets allocated the value e^{-1}. q controls the slope of the exponential function. The larger $|q|$ is chosen, and in case $q > 0$,

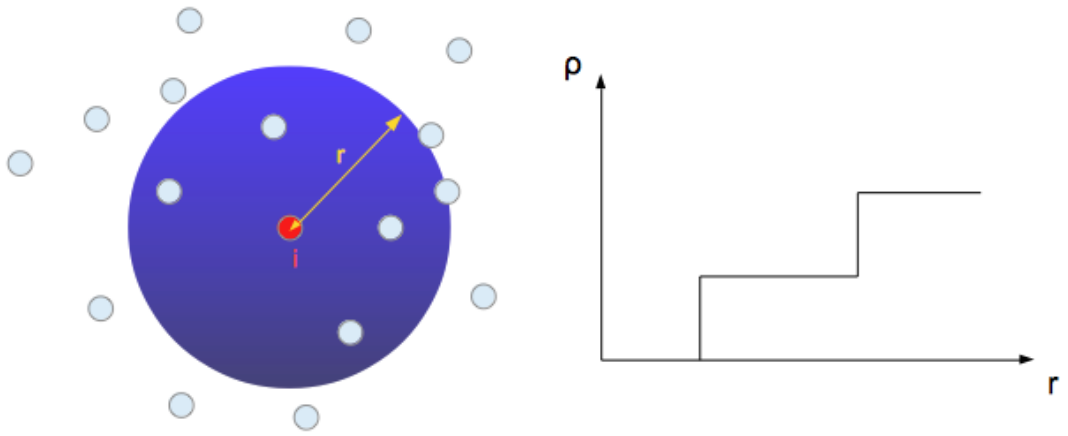

Figure 1.1: The number of neighbors of point i, $N_{neighbors,i}$ can be expressed by the cumulative distribution function $\rho_i(r)$, which "counts" the points in the vicinity of point i within a sphere of radius r (left). Point i is marked red. Neighboring points in the blue area are counted as 1, points in the white area not ("counted" as 0). Using a Θ-function as shaping function for $\rho_i(r)$ leads to a stepwise behavior of $\rho_i(r)$ (right, graph only qualitative). Therefore, $\rho_i(r)$ is not differentiable and a scaling range $[r_1, r_2]$ would be necessary for calculating $\alpha_i(r_1, r_2)$.

the more the exponential function gets similar to a Θ-function, and a nearby neighboring point is weighted stronger than a distant one. In case $q < 0$, the exponential function approaches the shape of an inverse Θ-function, and a distant neighboring point is weighted stronger than a nearby point. A figurative visualization is shown in figure (1.2): the former sharp edge of the sphere at radius r (figure 1.1) has become smooth.

The scaling index $\alpha_i(r)$ is the partial derivate of $\ln \rho_i(r)$ with respect to $\ln r$:

$$\alpha_i(r) = \frac{\partial \ln \rho_i(r)}{\partial \ln r} = \frac{\sum_{j=1}^{N} q \, (\frac{d_{ij}}{r})^q \, e^{-(\frac{d_{ij}}{r})^q}}{\sum_{j=1}^{N} e^{-(\frac{d_{ij}}{r})^q}} \tag{1.20}$$

1.4 Anisotropic scaling indices

In order to find anisotropies in a point distribution, you need to change the distance measure. So far, the Euklidean distance measure $\|\cdot\|_2$ was used where points in different directions contribute equally. In the following the distance measure $\|\cdot\|_A$ is used, with A being a metric tensor. In general, the transformation A can be split up into a rotational matrix R and a matrix M containing the eigenvalues of A: $A = R^T \cdot M \cdot R$. The eigenvalues will be denoted by λ_p, $p = 1, \ldots n$. The dimensionality of A equals n. In the first parts of this thesis, only the main directions along the axes x, y, z (for 3D distributions) or

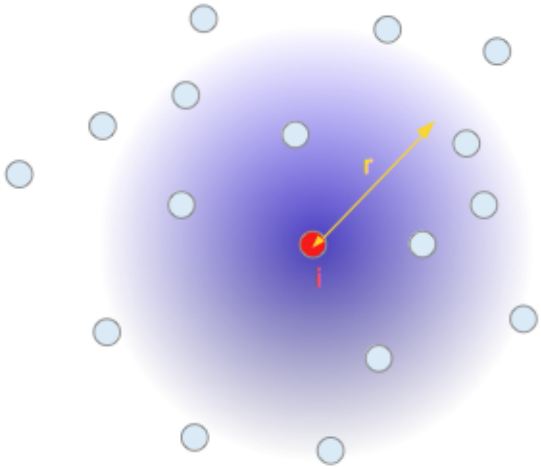

Figure 1.2: This figure sketches the shape of $\rho_i(r)$ using an exponential shaping function. Point i is marked red, the blue ones are neighboring points. The characteristic distance r marks the position where the exponential function attains the value e^{-1}. By choosing $q > 0$ nearby neighbors are weighted stronger than distant ones. By using a differentiable function as shaping function $s_r(d_{ij})$, the scaling index $\alpha_i(r)$ can be calculated without needing a scaling range $[r_1,\ r_2]$.

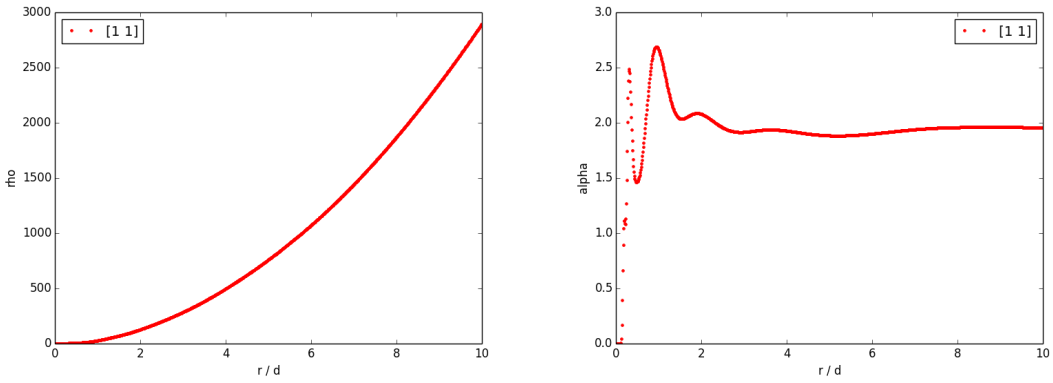

Figure 1.3: These graphs show the behavior of the weighted $\rho_i(r)$ (on the left) and the weighted $\alpha_i(r)$ (on the right) for an examplary point i. For large r (starting at about three average point distances d), $\alpha_i(r)$ attains the value of the dimensionality of the point distribution $n = 2$. The label [1 1] of the graphs means the isotropic Euklidean distance measure was used (see next section 1.4).

x, y (for 2D distributions) are considered. So the distance norm simplifies and the single eigenvalues $\lambda_p,\ p = 1, \ldots n$ control the strength of contribution of each dimension p to the total distance d_{ij}:

$$
\begin{aligned}
d_{ij} &= \left\{\vec{x}_{ij}^T \cdot A \cdot \vec{x}_{ij}\right\}^{\frac{1}{2}} = \left\{\vec{x}_{ij}^T \cdot R^T \cdot M \cdot R \cdot \vec{x}_{ij}\right\}^{\frac{1}{2}} \\
&= \left\{ \vec{x}_{ij}^T \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \vec{x}_{ij} \right\}^{\frac{1}{2}} \\
&= \left\{ \sum_{p=1}^{n} \lambda_p (x_{i,p} - x_{j,p})^2 \right\}^{\frac{1}{2}}
\end{aligned}
\tag{1.21}
$$

The ratio between larger and smaller λ_p is called aspect ratio, which in the following will be denoted by λ. The smaller λ_p in this thesis equals 1, the larger ones (in case of $n > 2$) equal each other. In case of $n = 3$, there are two ways of combining the eigenvalues which lead to different shapes of $\rho_i(r)$ (in this thesis denoted with "shape of $\rho_i(r)$" or just "shape"):

- The sets $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, \lambda, \lambda\},\ \{\lambda, 1, \lambda\},\ \{\lambda, \lambda, 1\},\ \lambda > 1$ lead to a cigar-like shape of $\rho_i(r)$ (illustration in figure 1.5).
- The sets $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, 1, \lambda\},\ \{1, \lambda, 1\},\ \{\lambda, 1, 1\},\ \lambda > 1$ lead to a discus-like shape of $\rho_i(r)$ (no illustration).

For $n = 2$, the eigenvalues set $\{\lambda_1, \lambda_2\} = \{1, \lambda\},\ \{\lambda, 1\},\ \lambda > 1$ is used. Figure (1.4) shows the consequential anisotropic shape of $\rho_i(r)$ in case of $n = 2$ for $\{\lambda_1, \lambda_2\} = \{\lambda, 1\},\ \lambda > 1$ (1st dimension being x-direction, 2nd dimension being y-direction).

If the calculation of $\alpha_i(r)$ along different spatial directions, i.e. for $\{\lambda_1, \lambda_2\} = \{1, \lambda\}$ and $\{\lambda_1, \lambda_2\} = \{\lambda, 1\},\ \lambda > 1$, gives different results, then there is an anisotropy in the point distribution, e.g. a linear structure (see chapter with examples).

1.5 The α-space

The analysis for any point i consists in the calculation of the scaling index α_i for one or more values of r (discrete steps in the algorithm) and for different (anisotropic) shapes and directions of ρ_i. The parameters q and λ are kept constant, once fixed. So overall one does not get a single scaling index for one point i, but a set of scaling indices $\{\alpha_k\}_i,\ k = 1, \ldots n_{par}$. The total number of parameters n_{par} results from the number of steps of r, n_r, the number of shapes, n_{shapes}, and the number of directions, n_{dir}: $n_{par} = n_r \cdot n_{shapes} \cdot n_{dir}$.

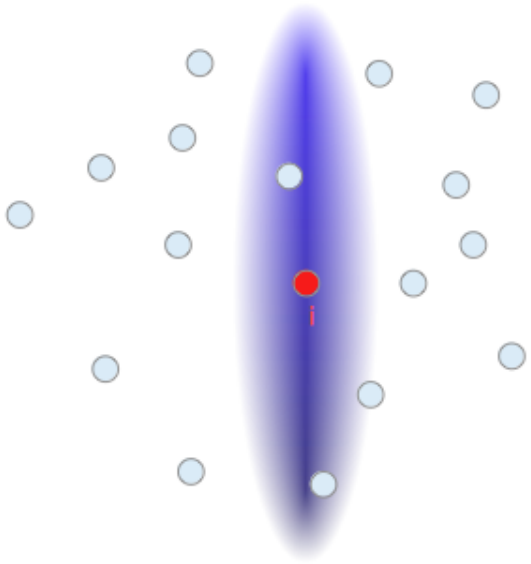

Figure 1.4: By using an anisotropic distance measure, $\rho_i(r)$ does not anymore weigh the neighboring points only according to their distance, but also according to their angular distribution around point i. Here the y-direction is preferred: the closer a neighboring point (blue) is distributed along the y-axis, the larger values it gets allocated (for $q > 0$).

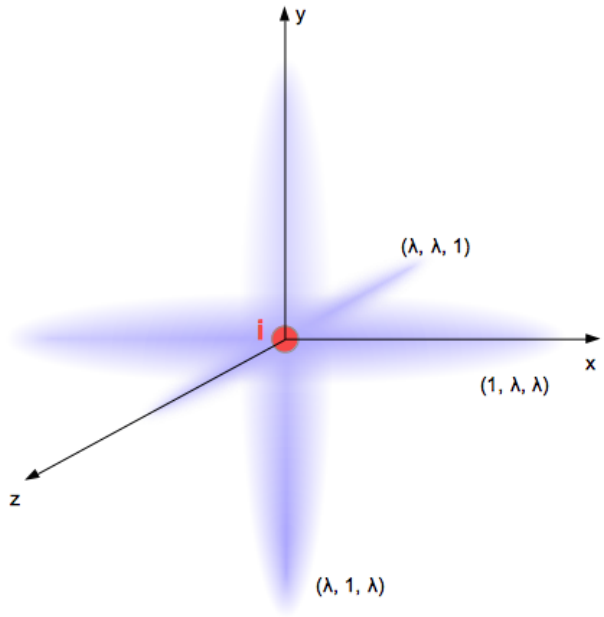

Figure 1.5: In case of $n = 3$ and using $\{\lambda_p\} = \{1, \lambda, \lambda\}, \ldots$ with $\lambda > 1$ the shape of $\rho_i(r)$ gets cigar-like.

The set $\{\alpha_k\}_i$ can be mapped onto a single point in the n_{par}-dimensional parameter-space, called α-space.

$$\mathbf{n_{par}} = \mathbf{n_r} \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \end{Bmatrix} \times \mathbf{n_{shapes}} \begin{Bmatrix} \text{cigar-like} \\ \text{discus-like} \end{Bmatrix} \times \mathbf{n_{dir}} \begin{Bmatrix} \text{direction 1 (e.g. x)} \\ \text{direction 2 (e.g. y)} \\ \text{direction 3 (e.g. z)} \\ \dots \end{Bmatrix} \tag{1.22}$$

As measure for the similarity of different point structures i, j the Euklidean-like distance Δ^{α}_{ij} is used. For larger distance in the α-space, $\Delta^{\alpha}_{i,j}$, the point structures $i,\ j$ are considered to be less similar. For reasons of smaller computing times, taking the square root was omitted:

$$\Delta^{\alpha}_{ij} = \sum_{k=1}^{n_{par}} (\alpha_{i,k} - \alpha_{j,k})^2 \tag{1.23}$$

1.6 Introduction to principal component analysis

Later in the thesis, it will be necessary to visualize the results. Due to its high dimensionality, the α-space first needs to be mapped into a 3D or 2D space. This is achieved by doing a principal component analysis PCA [21].

In a n-dimensional point distribution $\{\vec{x}_i\}, i = 1 \dots N$, the variance T of the distribution can also be expressed as sum over the variances T_p along the single dimensions $p = 1 \dots n$.

$$\begin{aligned} T(\{\vec{x}_i\}) &= \frac{1}{N} \sum_{i=1}^{N} \|\vec{x}_i - \vec{x}_{mean}\|^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{p=1}^{n} (x_{i,p} - x_{mean,p})^2 \\ &= \sum_{p=1}^{n} \frac{1}{N} \sum_{i=1}^{N} (x_{i,p} - x_{mean,p})^2 = \sum_{p=1}^{n} T_p(\{\vec{x}_i\}) \end{aligned} \tag{1.24}$$

The principal component analysis PCA [21] performs an orthogonal transformation $\vec{y}_i = V^T \cdot \vec{x}_i$ of the space vectors $\vec{x}_i$ in order to find a new basis with $m \leq n$ dimensions, where the first dimension $h = 1$ will contain the largest fraction of T: $T_{h=1}$; the second dimension $h = 2$ the second largest fraction, $T_{h=2}$; etc. The matrix V is composed of the eigenvectors $(\vec{v}_1\ \vec{v}_2\ \dots\ \vec{v}_m)$ of the correlation matrix of the dataset, where $\vec{v}_h$, $h = 1 \dots m$, is the eigenvector for eigenvalue μ_h, with $\mu_1 > \mu_2 > \dots > \mu_m$.

In this thesis, one aim of making a PCA is to be able to visualize the n_{par}-dimensional α-space in a 2D or 3D plot (so $m = 2$ or $m = 3$, respectively) without great loss of information, i.e. with keeping as much as possible of the total variance T. The first

dimension of the new basis, $h = 1$, which contains the largest fraction $T_{h=1} = \mu_1$ of the variance, will be denoted as the first principal component $\tilde{\alpha}_{h=1}$; the second dimension, $h = 2$, as second principal component $\tilde{\alpha}_{h=2}$; etc.

1.7 Introduction to bond order parameters

Bond order parameters [11] are a usual tool for analyzing structural properties of point distributions. They are local, i.e. only dependent on the property of the neighborhood of a local structure, and rotationally invariant, i.e. they are not dependent on the spatial orientation or scaling of a local structure. They have been used to analyze local structures in many systems, such as simple fluids, solids, and glasses [22] [11] [23] [24], hard-sphere systems [25] [26] [27] [28] [29] [30], colloidal suspensions [31] [32], 3D complex plasmas [7] [33] [34] [35] [36] and complex plasma films [37] [38].

"Bonds" in this context mean a kind of association between two points i, j (the following paragraphs follow [11]). In the following the next neighbors $j(i)$ of point i are considered to be bonds of i; the coordination number of the crystal types hcp and fcc is 12, and 8 for bcc (see section 3.1). The bond order parameters are the expansion coefficients Q_{lm} when expanding the bond distribution density $\rho(\theta, \phi)$ using the spherical harmonics Y_{lm}:

$$\rho(\theta, \phi) = \sum_{l,m} Q_{lm} \bar{Y}_{lm}(\theta(\vec{r}_{ij}), \phi(\vec{r}_{ij})) \tag{1.25}$$

θ, ϕ denote spherical coordinates. $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, with $\vec{r}_i$ and $\vec{r}_j$ being the position vectors of points i and j, respectively. Parameters l and m are degree and order of the Legendre polynomials, one has $l \geq 0$ and $-l \leq m \leq l$. So the bond order parameters are obtained by first calculating the spherical harmonics of the next neighbors $j(i)$ of point i. The spherical harmonics are denoted with Q_{lm} in the following.

$$Q_{lm}(\vec{r}_{ij}) \equiv Y_{lm}(\theta(\vec{r}_{ij}), \phi(\vec{r}_{ij})) \tag{1.26}$$

Rotationally invariant quantities are combinations of Q_{lm} of second and third order: Q_l and W_l (W_l is not used in this thesis):

$$Q_l(i) = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |\bar{Q}_{lm}(i)|^2 \right)^{\frac{1}{2}} \tag{1.27}$$

$$\bar{Q}_{lm}(i) = \frac{1}{N_{nn}(i)} \sum_{j(i) \in \{nn\}} Q_{lm}(\vec{r}_{ij}) \tag{1.28}$$

The abbreviation nn refers to the next neighbors of a point, so $N_{nn}(i)$ denotes the number of next neighbors of point i.

If considering only the invariants with even l, no direction has to be assigned to the bonds of point i, due to the invariance under inversion $\vec{r} \to -\vec{r}$ of the spherical harmonics

for even l: $Y_{lm}(\vec{r}) = (-1)^l\, Y_{lm}(-\vec{r})$. The first nonzero Q_l for cuboid crystal types such as fcc is Q_4. For hcp, Q_2 vanishes accidentally [11]. Therefore, the bond order parameters which are most easily calculated and normally used for the discrimination between crystalline structures are Q_4 and Q_6. Figure 1.6 shows a cut-out plot from Steinhardt et al. [11], which shows the bond order parameters $Q_l,\ l = 2, \ldots 10$, l even.

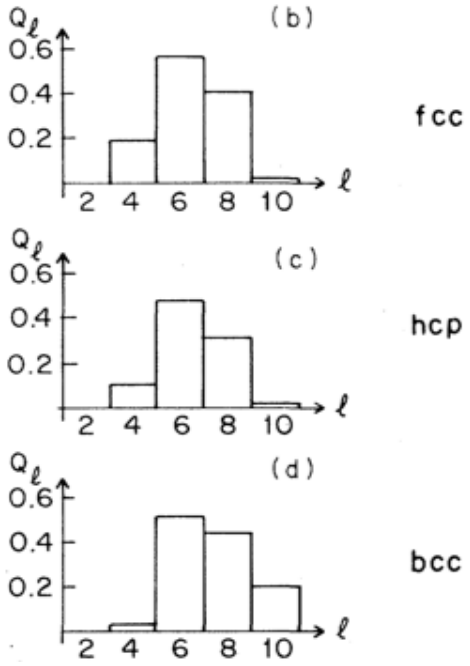

Figure 1.6: Cut-out from Steinhardt et al. [11]. The plot shows the signals of $Q_l,\ l = 2, 4, 6, 8, 10$ for crystal types fcc, hcp and bcc.

After having calculated the parameters for all points $i = 1 \ldots N$ of a distribution, they are allocated to a certain crystal type dependent on the distance in the $Q_4 - Q_6$ parameter space: $\Delta^{Q_4-Q_6}$. The crystal type t is chosen, which has the minimal distance $\min_{t=fcc,hcp} \Delta_{i,t}^{Q_4-Q_6}$. Index i refers to any point in the distribution. Additionally, the distance needs to be smaller than a certain threshold: $\Delta_{i,t}^{Q_4-Q_6} \overset{!}{<}$ threshold Bond.

Chapter 2

Descriptive examples with 2D point distributions

The usage of the scaling indices shall now be illustrated further with the aid of easy-structured two-dimensional point distributions. Within the following three distributions, some point i is chosen (which is marked red in the plots), whose isotropic and anisotropic scaling indices are calculated. As you will see, the anisotropic scaling indices will behave more or less equally in a more or less isotropic distribution, but show clear differences if the distribution contains anisotropic elements.

All calculations in this thesis are performed with own code written in Python.

2.1 Random point distribution

For an uniform random 2D point distribution ($N = 2245$ within the square $x, y \in [0, 3]$), I calculated the scaling index for a single point i. Figure (2.1) shows the point distribution, point i is marked red.

For calculating α_i, I used formula (1.20) with $q = 4$:

$$\alpha_i(r) = \frac{\sum_{j=1}^{N} 4 \, (\frac{d_{ij}}{r})^4 \, e^{-(\frac{d_{ij}}{r})^4}}{\sum_{j=1}^{N} e^{-(\frac{d_{ij}}{r})^4}} \tag{2.1}$$

The reasons for setting the value of the weighting parameter q, as well as the value of aspect ratio λ, are given in the next chapter, section 3.6. Radius r runs in the range $r \in [0.01d, 10.0d]$, step size $0.01d$. d is the estimated average point distance of N points within a volume V:

$$d = \sqrt[3]{\frac{V}{N}} \tag{2.2}$$

I made this calculation with the isotropic distance measure ($\lambda_1 = \lambda_2 = 1$), and with anisotropic distance measures ($\{\lambda_1, \lambda_2\} = \{1, 9\}$ and $\{\lambda_1, \lambda_2\} = \{9, 1\}$). The aspect ratio

9/1 will be denoted just by $\lambda = 9$ in the following. Figure (2.2) shows three graphs of $r - \alpha_i(r)$, with the characteristic distance r given in units of d (here $d \approx 0.159$), for isotropic and anisotropic distance measures with aspect ratio $\lambda = 9$.

Each graph $r - \alpha_i(r)$ shows four characteristic parts according to different values of r.
In the first part ($r \approx 0$) one has $\alpha_i(r) \approx 0$, because the distance d_{ij} of most points j is far larger than the characteristic distance r: $d_{ij} \gg r$ for most points j.
In the second part ($r \approx d$) more and more points in the neighborhood of i are counted by $\rho_i(r)$, so $\alpha_i(r)$ raises.
In the third part ($r \gg d$) $\alpha_i(r)$ saturates and reaches the value of the dimensionality of the point distribution.
In the fourth part (not always important and cut out in the plot), $\alpha_i(r)$ decreases again due to border effects, because this point distribution has a finite extension.

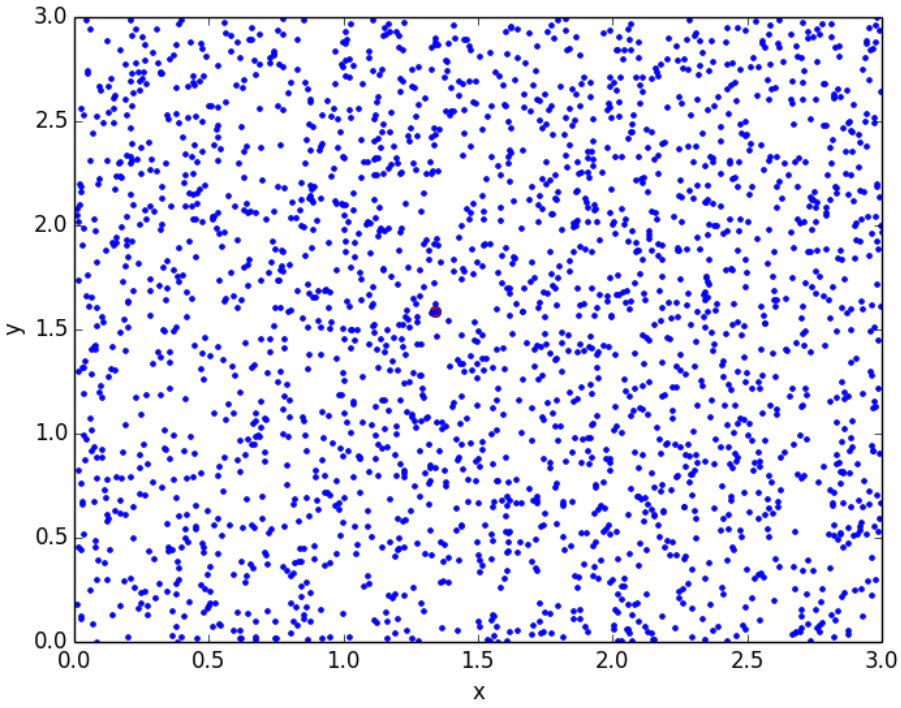

Figure 2.1: 2D random point distribution. The scaling index $\alpha_i(r)$ is calculated for the red marked point i in the middle, shown in figure (2.2).

2.2 Layer of a 3D fcc-like distribution

I repeated the analysis for a 2D point distribution which is symmetric regarding x- and y-direction (namely a single layer out of a 3D fcc-like point distribution) with equal size

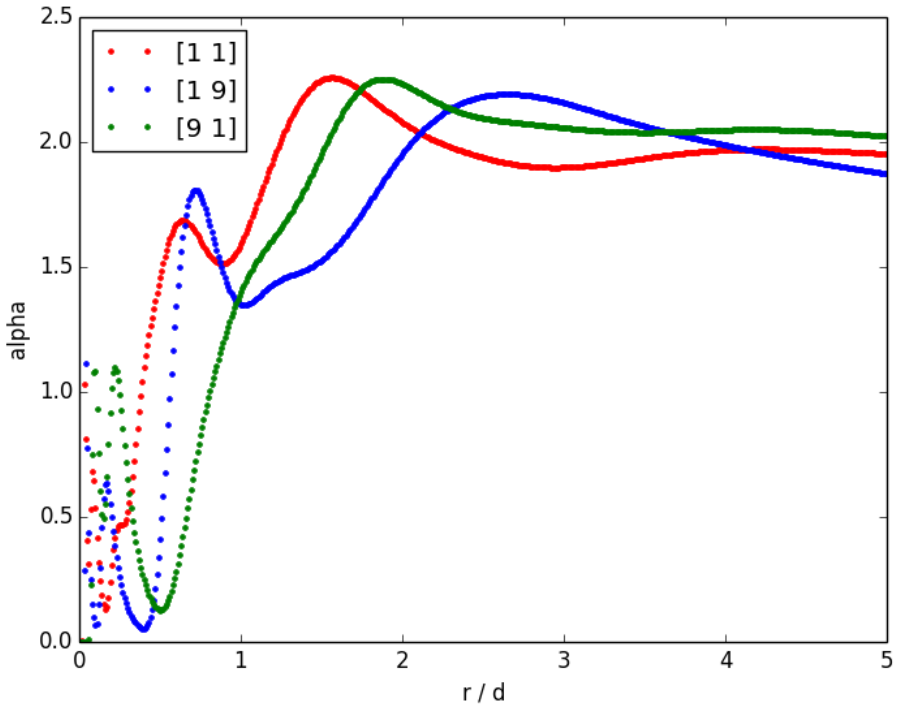

Figure 2.2: Graph of $r - \alpha_i(r)$ for point i in figure (2.1), for the isotropic distance measure $\lambda_1 = \lambda_2 = 1$ and anisotropic distance measures $\{\lambda_1, \lambda_2\} = \{1, 9\}$ and $\{\lambda_1, \lambda_2\} = \{9, 1\}$. r is given in units of the average point distance d. The two graphs for the anisotropic distance measures (blue and green) differ, because the random point distribution shows little anisotropies, as can be clearly seen in the vicinity of point i in figure (2.1). All three graphs saturate at $\alpha_i(r) \approx 2,\ r \gtrsim 3d$.

($N = 2245$ and $x, y \in [0, 3]$). Figure (2.3) shows the point distribution, point i is marked red. The resulting graphs of $r - \alpha_i(r)$ are shown in figure (2.4). The graphs of $r - \alpha_i(r)$ for $\{\lambda_1, \lambda_2\} = \{1, 9\}$ (blue) and $\{\lambda_1, \lambda_2\} = \{9, 1\}$ (green) are identical, meaning the fcc layer has an isotropic structure along these directions.

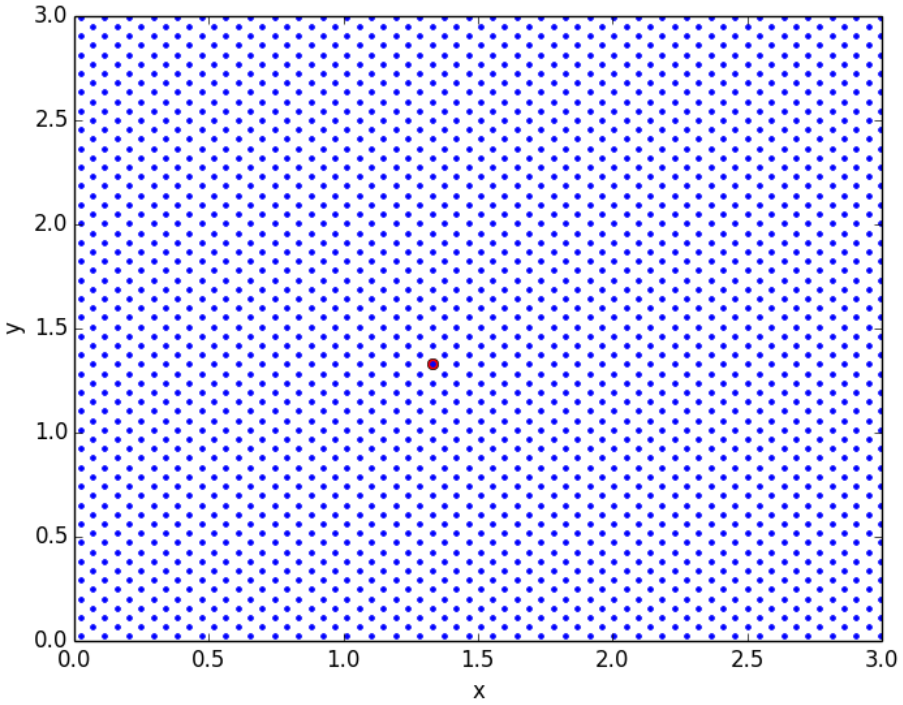

Figure 2.3: 2D layer out of a 3D fcc-like distribution. The layer is symmetric regarding x- and y-direction. The scaling index $\alpha_i(r)$ is calculated for the red marked point i in the middle, shown in figure (2.4).

2.3 Point distribution containing linear structures

In a random point distribution (similar, but not equal to the one above, $N = 2245$ and $x, y \in [0, 3]$) the x-dimension was rounded to the first decimal, which leads to vertical (along y-direction) linear structures in the distribution. Figure (2.5) shows the point distribution, point i is marked red. The analyses made with isotropic and anisotropic distance measures are shown in figure (2.6). The difference between the distance measures with $\{\lambda_1, \lambda_2\} = \{1, 9\}$ and with $\{\lambda_1, \lambda_2\} = \{9, 1\}$ can be seen clearly. For $\{\lambda_1, \lambda_2\} = \{9, 1\}$, $\alpha_i(r)$ first reaches a value of about 1, meaning that the point distribution behaves one-dimensional at the beginning for small r, because only points of one linear structure are counted by $\rho_i(r)$ for small r. Later on, for larger r, also the other lines have influence on the scaling index, and $\alpha_i(r)$ reaches 2 according to a 2D point distribution.

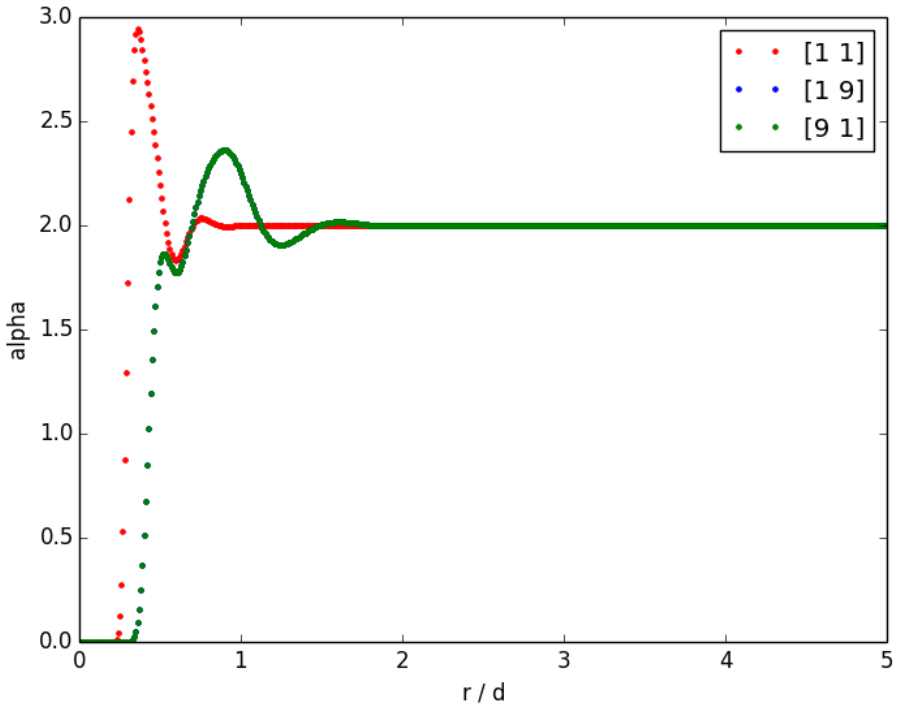

Figure 2.4: Graph of $r - \alpha_i(r)$ of point i in figure (2.3), for the isotropic distance measure $\lambda_1 = \lambda_2 = 1$ and anisotropic distance measures $\{\lambda_1, \lambda_2\} = \{1, 9\}$ and $\{\lambda_1, \lambda_2\} = \{9, 1\}$. r is given in units of the average point distance d. The blue and green graphs lie one on top of the other, meaning the point distribution is isotropic along the measured directions (x- and y-direction). At $r \approx 2d$, $\alpha_i(r)$ has saturated.

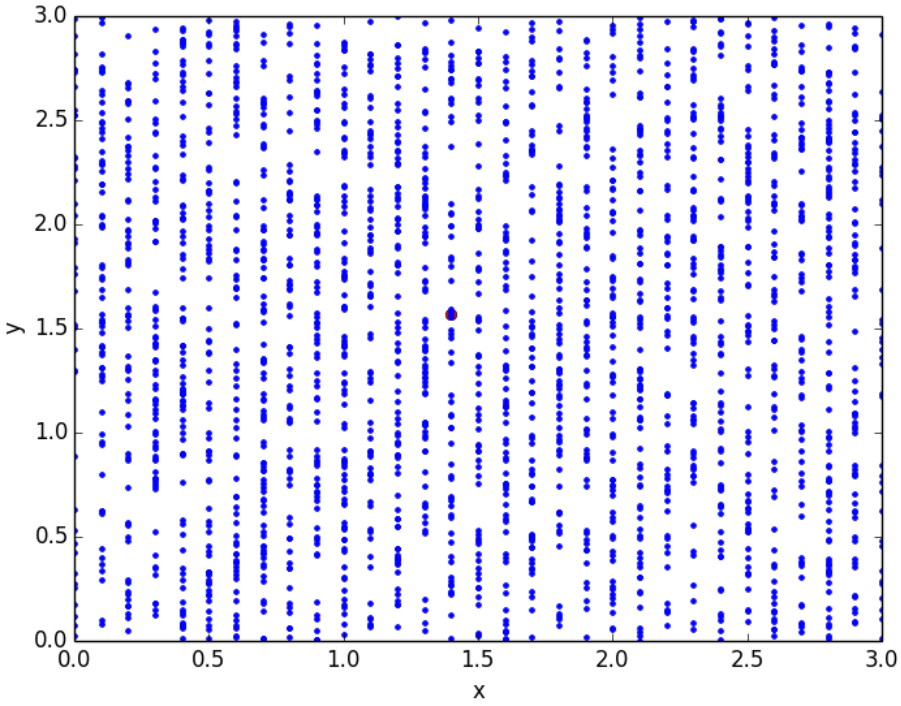

Figure 2.5: 2D random point distribution, with x-dimension been rounded to first decimal, leading to vertical lines. The scaling index $\alpha_i(r)$ is calculated for the red marked point i in the middle, shown in figure (2.6).

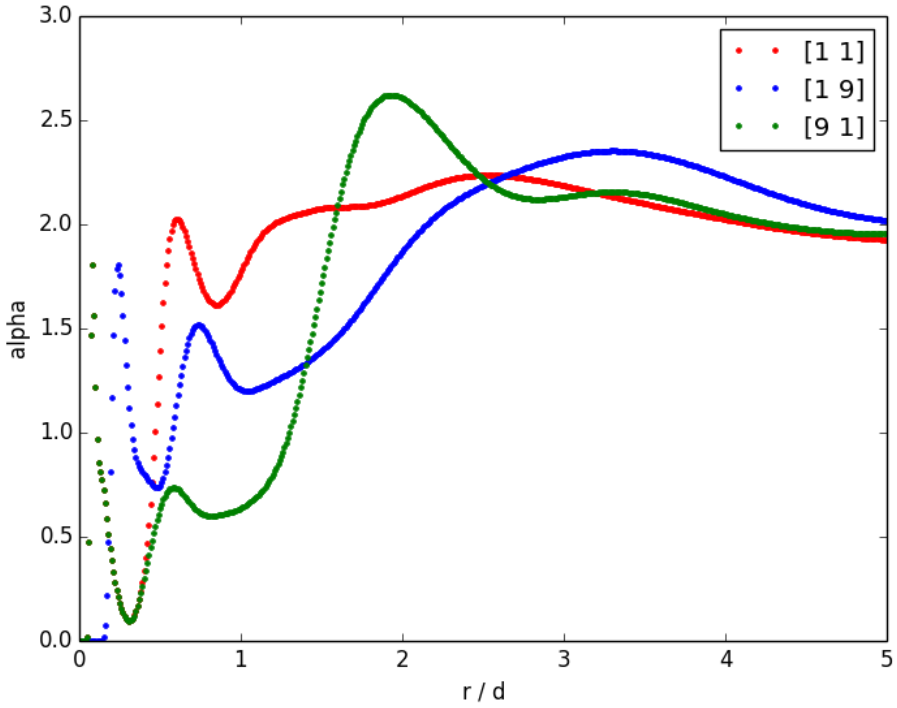

Figure 2.6: Graph of $r - \alpha_i(r)$ of point i in figure (2.5), for the isotropic distance measure $\lambda_1 = \lambda_2 = 1$ and anisotropic distance measures $\{\lambda_1, \lambda_2\} = \{1, 9\}$ and $\{\lambda_1, \lambda_2\} = \{9, 1\}$. r is given in units of the average point distance d. The graphs for the anisotropic distance measures differ clearly. For $\{\lambda_1, \lambda_2\} = \{9, 1\}$ (green), $\alpha_i(r)$ first reaches a value of about 1, meaning that the point distribution behaves one-dimensional at the beginning for small r, because only points of one linear structure are counted by $\rho_i(r)$ for small r. Later on, for larger r, also the other lines have influence on the scaling index, and $\alpha_i(r)$ reaches 2 according to a 2D point distribution. For the eigenvalues $\{\lambda_1, \lambda_2\} = \{1, 9\}$ (blue), the other lines have impact on $\alpha_i(r)$ quite early (for small r), due to the horizontally aligned shape of $\rho_i(r)$, why $\alpha_i(r)$ increases to larger values earlier.

Chapter 3

Quality of discrimination between crystal types: comparing scaling indices and bond order parameters

The usability of the scaling indices for the purpose of discrimination of several crystal types in a point distribution is tested in the following chapter. The point distributions are analyzed with the method of bond order parameters, too. Then the results are compared to each other with respect to the maximal noise level of the point distributions, at which the crystal types still can be differed.

3.1 Description of the datasets used in this chapter

The crystal types which shall be discriminated are namely bcc ("body-centered cubic"), fcc ("face-centered cubic "or "cubic close packed",) and hcp ("hexagonal close packed"). The figures (3.1) show schematic plots of the unit cells. The figures (3.2), (3.3) and (3.4) show the artificially created point distributions which are used in this thesis. They are originally of perfect shape, with additive Gaussian noise (with its standard deviation given in percentage of the average point distance). Cut-outs showing the crystal structure are shown in the appendix in figures (7.1, 7.2, 7.3).

The crystal types bcc and the close-packed fcc are cubical bravais lattices. By "Bravais lattices", the possible periodic lattices are classified. In three dimensions, there are 14 Bravis lattices, each with certain symmetries. Bcc and fcc are cubical (with equal side lengths a and rectangular axes), they have symmetry of rotations by 90° around an axis, symmetry of inversion ($\vec{r} \to -\vec{r}$), and symmetry of reflection at certain planes. Bcc has a coordination number (number of next neighbors for any point) of 8; fcc one of 12. Hcp is no bravais lattice. It is formed by putting two hexagonal bravais lattices together (see figure 3.1): between two hexagons, which have distance c, sidelength a and each contain seven points, there are three more points. For close-packing, $c = \sqrt{\frac{8}{3}}a$ needs to be fulfilled.

The crystal types fcc and hcp are both close-packed types. Close-packing means following: if stacking balls in a box of volume V, as much volume as possible is occupied, $V_{occupied} < V$. This is achieved if the balls of the first layer are put together in hexagonal way (stacking order A), and the balls of the second layer in the hollows having formed inbetween the balls of the first layer (see figure 7.4). For doing this, there are two possible positions (stacking order B or C). Fcc is formed for repeating the stacking orders $ABCABC\ldots$, hcp is formed for $ABAB\ldots$.

Three artificially created 3D point distributions of bcc-, fcc- or hcp-like structure with different noise levels were analyzed. The orientations of the distributions are arbitrary, but are fixed. The number of points contained and the spatial extensions are:

- bcc-dataset: $N_{bcc} = 3456$ points in the boundaries $x, y, z \in [0.0, 1.1]$
- fcc-dataset: $N_{fcc} = 2916$ points in the boundaries $x, y, z \in [0.0, 0.8]$
- hcp-dataset: $N_{hcp} = 2820$ points in the boundaries $x, y, z \in [0.0, 0.13]$

The reason for the different boundaries is to achieve a similar number of points for all datasets. The point distributions are originally of perfect structure, with additive Gaussian noise. In the first part of this chapter its standard deviation was either 5% or 10% of the estimated average point distance, in the following called (a) or (b), respectively.

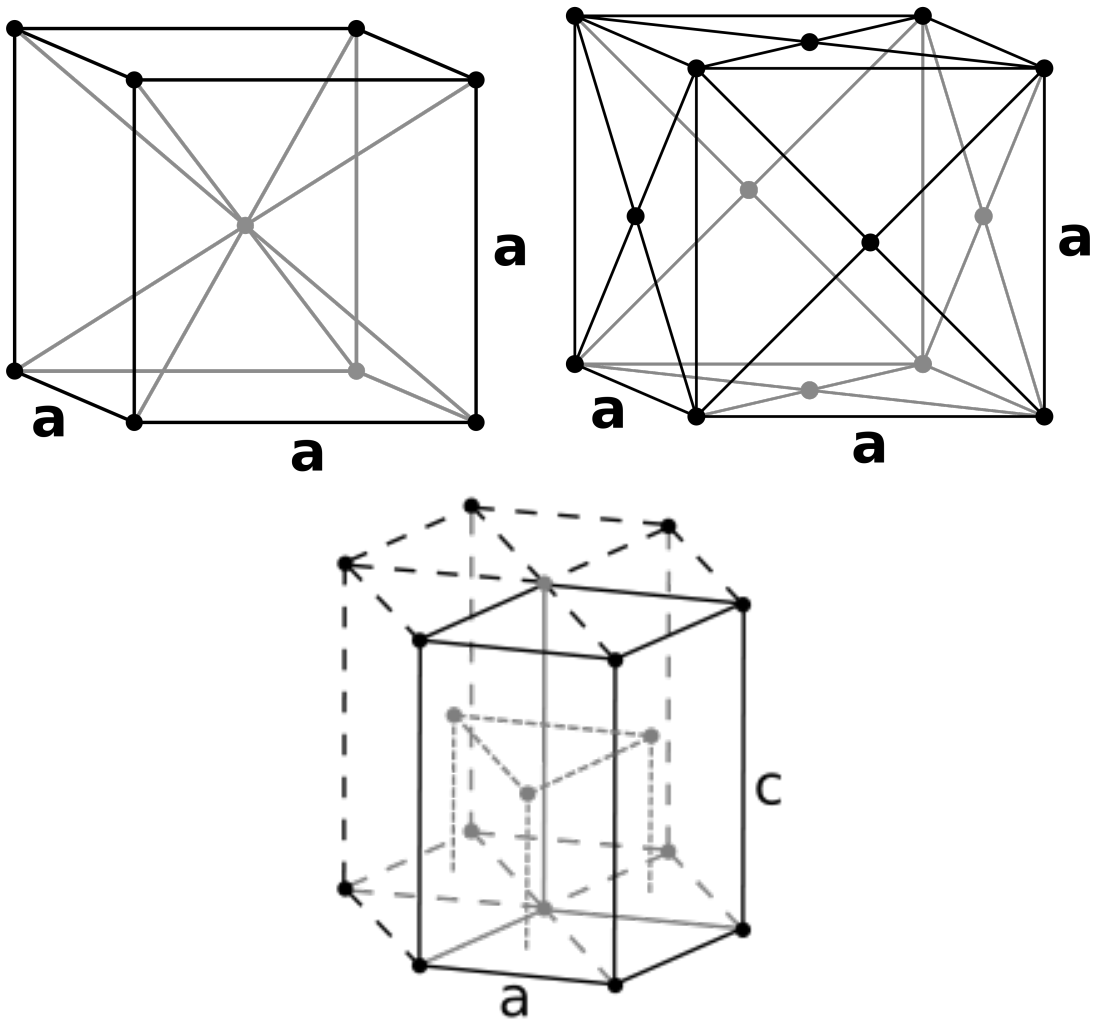

Figure 3.1: Schematic plots of the unit cells of the crystal types bcc (top left), fcc (top right) and hcp (bottom) ([39] [40] [41]). The unit cell of bcc contains 8 points in the corners and 1 in the middle of its body. The coordination number is 8. The side lengths are equal, denoted with a.
The unit cell of fcc contains 8 points in the corners and 6 in the middles of its facets. The coordination number is 12. The side lengths are equal, denoted with a.
The unit cell of hcp is built by two hexagons each containing 6 points in the corners and 1 in the middle. Their side length a and distance c has the ratio $\frac{c}{a} = \sqrt{\frac{8}{3}}$. Inbetween there are three more points. The hcp type is characterized by this stacking order $ABAB\ldots$ (see figure 7.4 appendix). The coordination number is 12.

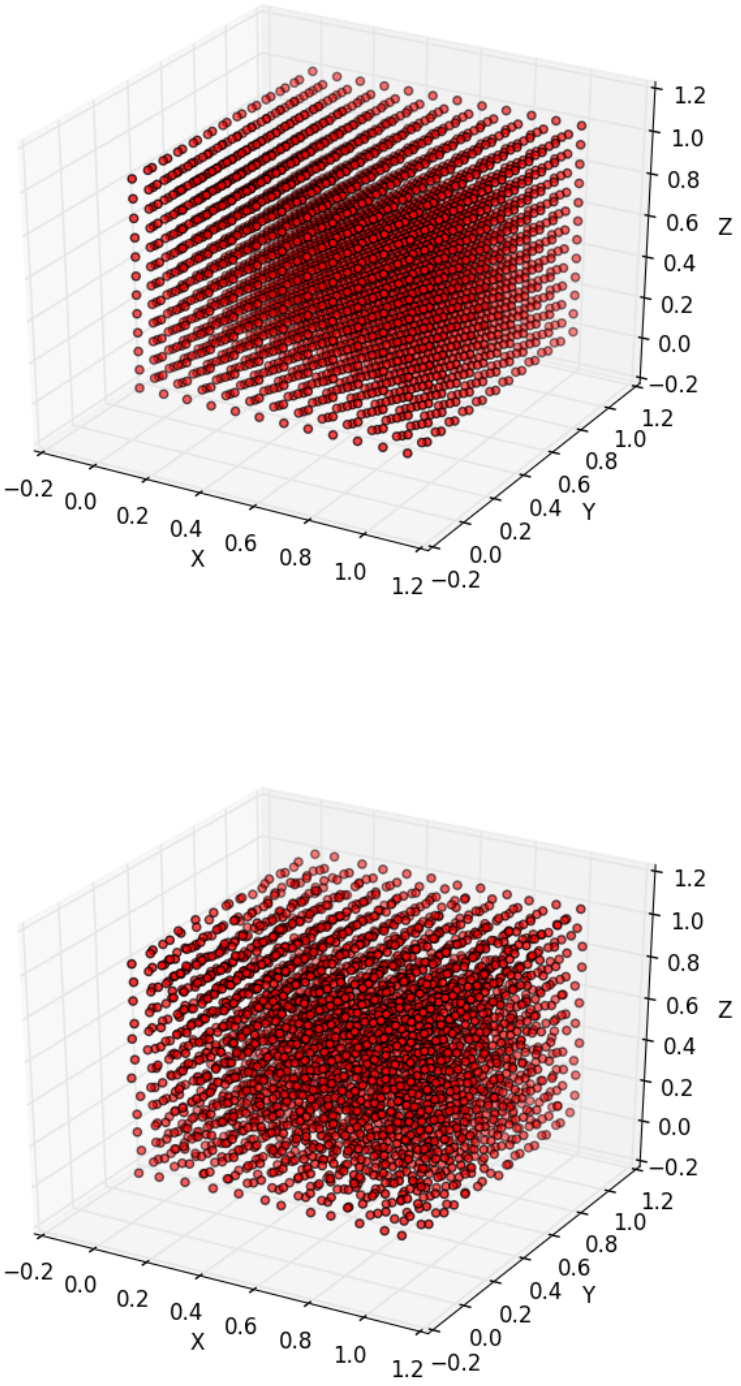

Figure 3.2: Artificially created point distribution of bcc type. The noise levels (standard deviation) of the additive Gaussian noise are 1% (top) of the average point distance and 10% (bottom), respectively.

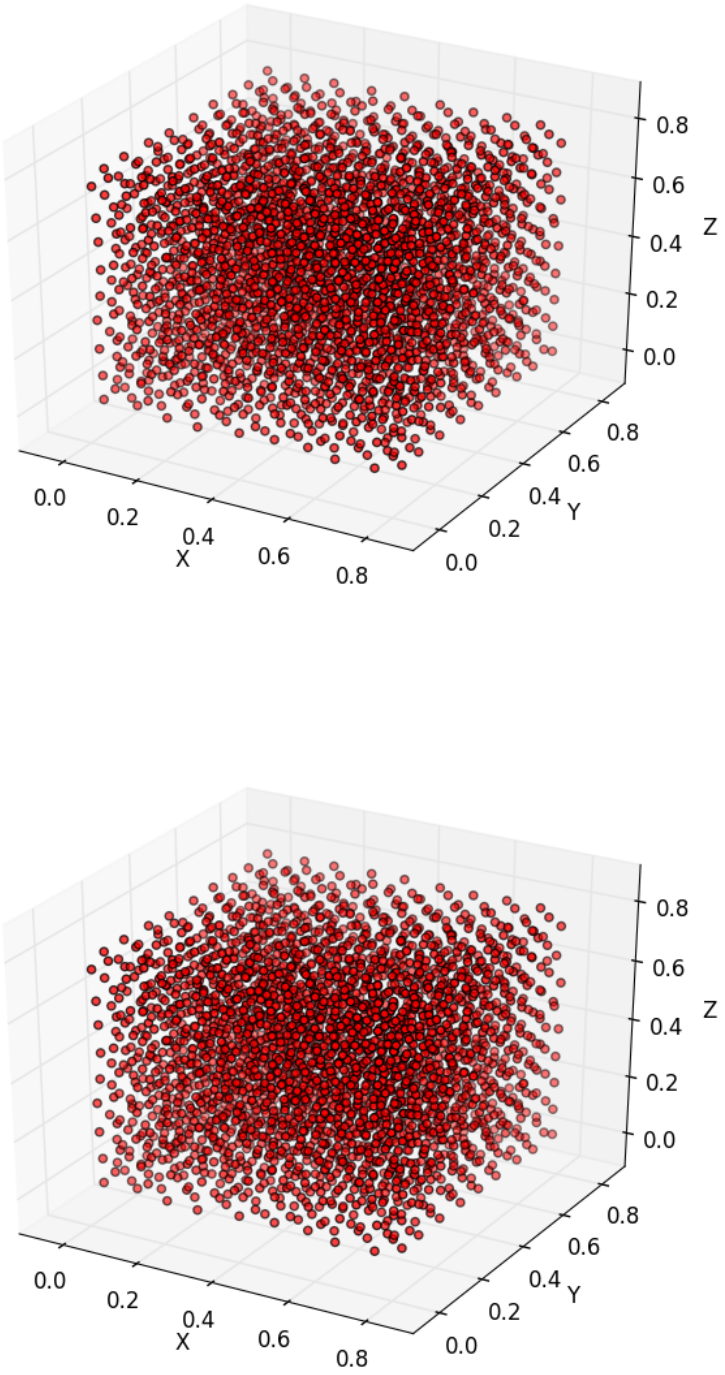

Figure 3.3: Artificially created point distribution of fcc type. The noise levels (standard deviation) of the additive Gaussian noise are 1% (top) of the average point distance and 10% (bottom), respectively.

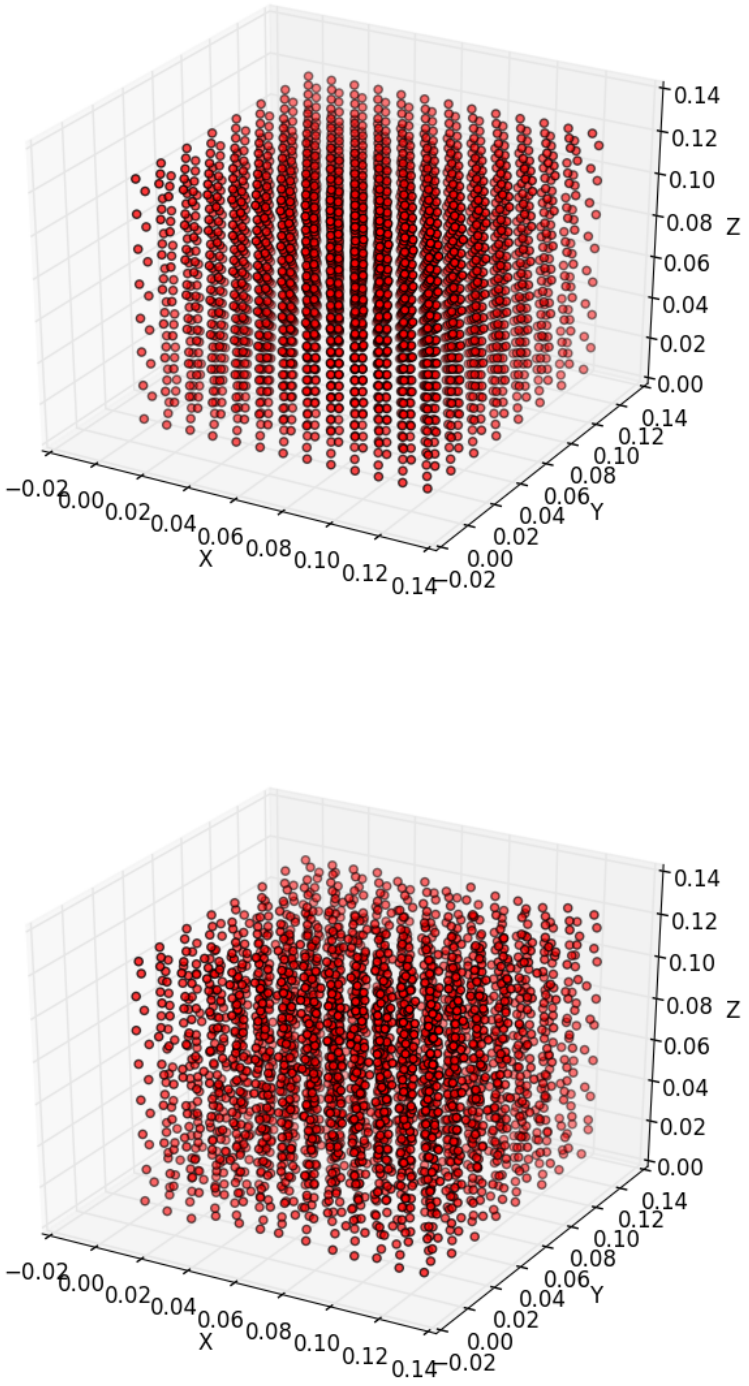

Figure 3.4: Artificially created point distribution of hcp type. The noise levels (standard deviation) of the additive Gaussian noise are 1% (top) of the average point distance and 10% (bottom), respectively.

3.2 Details about the analysis

For each point distribution of bcc, fcc and hcp type, a scaling indices analysis was performed for the middle of the point distribution for excluding edge effects. The boundaries of these middle regions are as follows; the number of inner points fluctuates a bit due to different noise, it is approximately 330 - 360.

- bcc-dataset: $x, y, z \in [0.3, 0.8]$
- fcc-dataset: $x, y, z \in [0.2, 0.6]$
- hcp-dataset: $x, y, z \in [0.03, 0.09]$

The analyses were done for different shapes of $\rho_i(r)$ by using certain eigenvalues of the distance measure, $\{\lambda_1, \lambda_2, \lambda_3\}$:

- $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, \lambda, \lambda\}, \{\lambda, 1, \lambda\}, \{\lambda, \lambda, 1\}$ yields a cigar-like shape of $\rho_i(r)$
- $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, 1, \lambda\}, \{1, \lambda, 1\}, \{\lambda, 1, 1\}$ yields a discus-like shape of $\rho_i(r)$
- An isotropic distance measure turned out to be not adequate for the discrimination purpose due to poor results.

If the analysis includes both shapes, the set $\{\alpha_k\}_i,\ k = 1, \ldots n_{par}$ contains six scaling indices: $n_{par} = n_r \cdot n_{shapes} \cdot n_{dir} = 1 \cdot 2 \cdot 3 = 6$, i.e. each data point i is mapped onto a point in a six-dimensional α-space. Doing this for all three point distributions of bcc, fcc and hcp, one gets three different and ideally easily distinguishable α_i-distributions in the α-space. If only one shape is used, $n_{shapes} = 1$, one gets resulting α_i-distributions in a 3-dimensional α-space. In order to visualize these α_i-distributions, I made a principal component analysis which maps the α_i-distributions in the six-dimensional or three-dimensional α-space onto a two-dimensional plane.

The variances of the resulting α_i-distributions for each bcc-, fcc- or hcp-like point distribution are denoted with $T_{type},\ type = bcc, fcc, hcp$. The total variance of the three α_i-distributions together is denoted with T_{all}. The larger the ratio of variances $T_{all} / \max(T_{type})$, the better the crystal types can be discriminated. The values of the parameters $\lambda,\ q,\ r$ were kept constant and chosen in such a way, that the discrimination between the distributions of bcc, fcc and hcp type could be done best, i.e. for maximal ratio of variances $T_{all} / \max(T_{type}),\ type = bcc, fcc, hcp$, see section 3.6. For this purpose several series of tests were performed: in the first series the quality of discrimination was tested for different λ without specifying q and r. After having fixed λ, in the second series q was fixed without specifying r. Finally, in the third series r was fixed. The parameters chosen were $\lambda = 9,\ q = 4,\ r = 1.8\ d$.

3.3 Analysis with cigar- and discus-like shapes

In this test I calculated the scaling indices of the data set using both shapes of $\rho_i(r)$: the cigar-like shape of $\rho_i(r)$ by using $\{\lambda_1,\ \lambda_2,\ \lambda_3 = \{1, 9, 9\}, \{9, 1, 9\}, \{9, 9, 1\}$ and the discus-like shape of $\rho_i(r)$ by using $\{\lambda_1,\ \lambda_2,\ \lambda_3\} = \{1, 1, 9\}, \{1, 9, 1\}, \{9, 1, 1\}$. Figure (3.5) shows the result of the analyses for both noise-levels (a) and (b). The x-axis shows the first principal component $\tilde{\alpha}_{h=1}$ of the 6-dimensional α-space, the y-axis the second one $\tilde{\alpha}_{h=2}$.

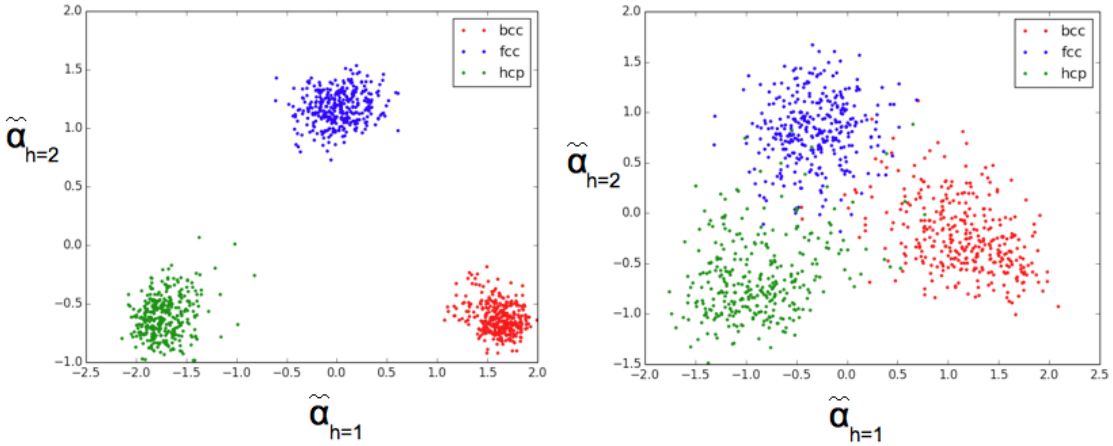

Figure 3.5: The plot shows a PCA of the scaling indices analysis for the datasets (a) on the left, and (b) on the right. Both cigar-like and discus-like shapes are being used ($n_{par} = n_r \cdot n_{shapes} \cdot n_{dir} = 1 \cdot 2 \cdot 3 = 6$). The additive Gaussian noise has a standard deviation of 5% (left) and 10% (right) of the average point distance. The first principal component $\tilde{\alpha}_{h=1}$ of the six-dimensional α-space is given on the x-axis, the second principal component $\tilde{\alpha}_{h=2}$ on the y-axis.

Results for the datasets (a):

- Total variance $T_{all} \approx 2.69$
- Variances $T_{all,h}$ along the principal components $h = 1,\ 2$:
  $T_{all,h=1} = \mu_1 \approx 1.86, T_{all,h=2} = \mu_2 \approx 0.74$
- Transformation matrix: $V^T \approx \begin{pmatrix} 0.05 & -0.72 & -0.63 & 0.25 & -0.01 & -0.10 \\ 0.69 & -0.01 & 0.08 & -0.20 & -0.46 & -0.52 \end{pmatrix}$
- Ratio of variances $\frac{T_{all}}{\max_{type=bcc,fcc,hcp}(T_{type})} \approx 15.2$

Results for the datasets (b):

- Total variance $T_{all} \approx 1.74$
- Variances $T_{all,h}$ along the principal components $h = 1,\ 2$
  $T_{all,h=1} = \mu_1 \approx 0.88, T_{all,h=2} = \mu_2 \approx 0.51$

- Transformation matrix: $V^T \approx \begin{pmatrix} -0.04 & -0.70 & -0.66 & 0.25 & 0.06 & -0.03 \\ 0.88 & -0.08 & 0.04 & 0.06 & -0.31 & -0.36 \end{pmatrix}$
- Ratio of variances $\frac{T_{all}}{\max_{type=bcc,fcc,hcp}(T_{type})} \approx 2.2$

The α_i-distributions are clearly distinguishable in case (a), the ratio of variances is about 15. In case (b), where the ratio of variances barely exceeds 2, the discrimination is just sufficient. So the scaling indices method is usable up to a noise level of 10% of the average point distance. The transformation matrix V yields that the cigar-like shape has a larger share in the first principal component $\tilde{\alpha}_{h=1}$ than the discus-like shape. The discus-like shape has dominant share in the second principal component $\tilde{\alpha}_{h=2}$.

3.4 Analysis with only cigar-like shape

In this section I calculated the scaling indices of the data set using only the cigar-like shape, $\{\lambda_1, \lambda_2, \lambda_3 = \{1, 9, 9\}, \{9, 1, 9\}, \{9, 9, 1\}$. Figure (3.6) shows the result of the analyses for both noise-levels (a) and (b). The x-axis shows the first principal component $\tilde{\alpha}_{h=1}$ of the 3-dimensional α-space, the y-axis the second one $\tilde{\alpha}_{h=2}$.

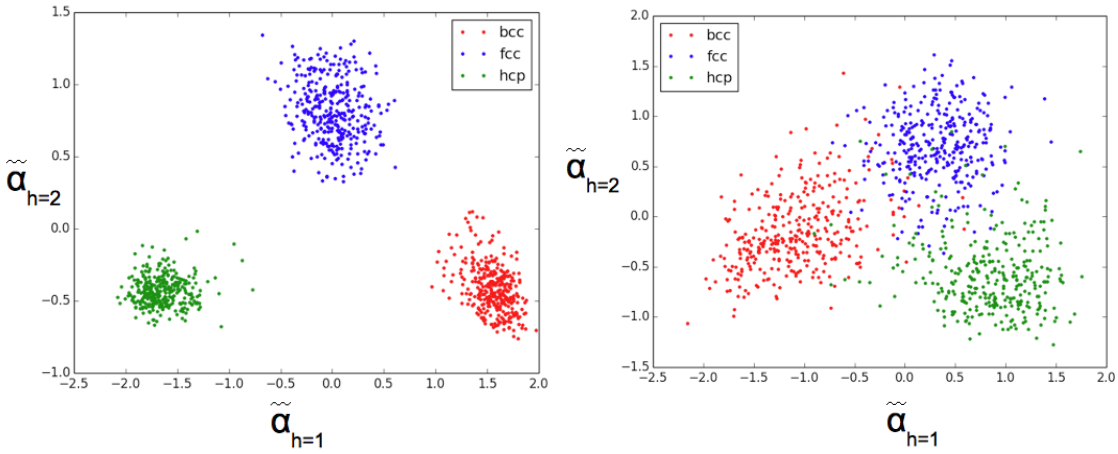

Figure 3.6: The plot shows a PCA of the scaling indices analysis for the dataset (a) on the left, and (b) on the right. Only the cigar-like shape is being used here ($n_{par} = n_r \cdot n_{shapes} \cdot n_{dir} = 1 \cdot 1 \cdot 3 = 3$). The additive Gaussian noise has a standard deviation of 5% (left) and 10% (right) of the average point distance. The first principal component $\tilde{\alpha}_{h=1}$ of the three-dimensional α-space is given on the x-axis, the second principal component $\tilde{\alpha}_{h=2}$ on the y-axis.

Results for the datasets (a):

- Total variance $T_{all} \approx 2.12$

- Variances $T_{all,h}$ along the principal components $h = 1,\ 2$
  $T_{all,h=1} = \mu_1 \approx 1.72, T_{all,h=2} = \mu_2 \approx 0.37$
- Transformation matrix: $V^T \approx \begin{pmatrix} 0.05 & -0.75 & -0.66 \\ 0.99 & -0.03 & 0.11 \end{pmatrix}$
- Ratio of variances $\frac{T_{all}}{\max_{type=bcc,fcc,hcp}(T_{type})} \approx 15.8$

Results for the datasets (b):

- Total variance $T_{all} \approx 1.37$
- Variances $T_{all,h}$ along the principal components $h = 1,\ 2$
  $T_{all,h=1} = \mu_1 \approx 0.83, T_{all,h=2} = \mu_2 \approx 0.41$
- Transformation matrix: $V^T \approx \begin{pmatrix} 0.06 & 0.72 & 0.69 \\ 0.99 & -0.12 & 0.03 \end{pmatrix}$
- Ratio of variances $\frac{T_{all}}{\max_{type=bcc,fcc,hcp}(T_{type})} \approx 2.7$

This analysis only using the cigar-like shape of $\rho_i(r)$ brings no worse result than the analysis using both geometries; however, it only needs the half time. So, there is no need to use also the discus-like shape of $\rho_i(r)$ so far.

3.5 Limit noise level for sufficient discrimination

Now the limit noise level shall be determined, for which the discrimination of the crystal types still works. For doing this, the bcc-, fcc- and hcp-like point distributions were analyzed for different Gaussian noise levels 1% $d \ldots 50\%$ d (d is the average point distance). The parameters of the analyses are $n_{shapes} = 2,\ n_{dir} = 3,\ n_r = 1,\ \{\lambda, q, r\} = \{9, 4, 1.8\}$.

The limit noise level is defined by the limit ratio of variances:

$$\frac{T_{all}(\text{limit noise level})}{\max_{type=bcc,fcc,hcp}(T_{type}(\text{limit noise level}))} = 2 \tag{3.1}$$

This limit ratio of variances was defined by means of a suitable cost-benefit-ratio between right and wrong allocations. Assume two n-dimensional point distributions $\{\vec{x}_i\}_i,\ i = 1, \ldots N_1$ and $\{\vec{x}_j\}_j,\ j = 1, \ldots N_2$ with centers at $\vec{\mu}_1$ or $\vec{\mu}_2$, respectively, and with equal variance σ^2 (equal variance is assumed due to equal noise levels of all structures during the same measurement). The total number of points WLOG is denoted with $N = N_1 + N_2 = \beta_1 N + \beta_2 N$. When joining the distributions, $\{\vec{x}_k\}_k = \{\vec{x}_i\}_i \cup \{\vec{x}_j\}_j,\ k = 1, \ldots N$, their center shall be at 0; this is fulfilled for $N_1\vec{\mu}_1 + N_2\vec{\mu}_2 = 0$. The total variance T_{all} then is:

$$\begin{aligned} T_{all} &= \frac{1}{N}\sum_{k=1}^{N}||\vec{x}_k - \vec{0}||^2 = \frac{1}{N}\left[\sum_i ||\vec{x}_i||^2 + \sum_j ||\vec{x}_j||^2\right] \\ &= \frac{1}{N}\left[\sum_{i,p} x_{i,p}^2 + \sum_{j,p} x_{j,p}^2\right] \end{aligned} \tag{3.2}$$

Completing the squares then yields:

$$\begin{aligned} T_{all} &= \frac{1}{N}\sum_{i,p}\left[(x_{i,p}-\mu_{1,p})^2 - \mu_{1,p}^2 + 2x_{i,p}\mu_{1,p}\right] + \dots \\ &= \frac{\beta_1}{N_1}\left[\sum_{i,p}(x_{i,p}-\mu_{1,p})^2\right] - \frac{\beta_1}{N_1}\left[\sum_{i,p}\mu_{1,p}^2\right] + \frac{\beta_1}{N_1}\left[\sum_{i,p}[2x_{i,p}\mu_{1,p}]\right] + \dots \\ &= \beta_1\sigma^2 - \beta_1||\vec{\mu}_1||^2 + 2\beta_1||\vec{\mu}_1||^2 + \dots \\ &= \beta_1\sigma^2 + \beta_1||\vec{\mu}_1||^2 + \beta_2\sigma^2 + \beta_2||\vec{\mu}_2||^2 \end{aligned} \tag{3.3}$$

Assuming symmetric distributions, $\vec{\mu}_1 = -\vec{\mu}_2,\ ||\vec{\mu}_1|| =: \mu$:

$$T_{all} = \sigma^2 + \mu^2 \tag{3.4}$$

So if "touching" is defined as $||\vec{\mu}_1 - \vec{\mu}_2|| = 2\mu \stackrel{!}{=} 2\sigma$:

$$T_{all} = \sigma^2 + \sigma^2 = 2\sigma^2 \quad \rightarrow \frac{T_{all}}{\sigma^2} = 2 \tag{3.5}$$

Assuming the touching distributions are normally distributed, about 68% of their points have a distance to their original center of maximal σ, and therefore would be allocated right. I considered this cost-benefit-ratio as acceptable and defined the limit ratio of variances as 2 in general.

Figure 3.7 shows the results, the horizontal red line marks the ratio of variances equaling two. Between the noise levels 10% d and 11% d, the ratio of variances falls below two. Itemized by pairs, the limit noise level for bcc – fcc is 11% d; for bcc – hcp it is 10% d; for fcc – hcp it is 9% d.

For comparison reasons, I made an analysis of the same datasets with the method of bond order parameters, too. The resulting limit noise level for a good discrimination between bcc, fcc and hcp is 7% d. Itemized by pairs it is 5% d for a discrimination between bcc and hcp, 7% d for a discrimination between fcc and hcp and 8% d for a discrimination between bcc and fcc. The direct comparison is shown in the figure (3.8) for bcc, fcc and hcp together, and in the figures (3.9) for the pairs. For reasons of visualization of the $Q_4 - Q_6$ parameter space and the meaning of the ratio of variances, fig (3.10) shows the resulting $Q_4 - Q_6$ parameter space for the point distributions of bcc, fcc and hcp with a noise level of 5% d (on the left) or 10% d (on the right), respectively. The clouds in the $Q_4 - Q_6$ space begin to overlap once the ratio of variances falls below 2.

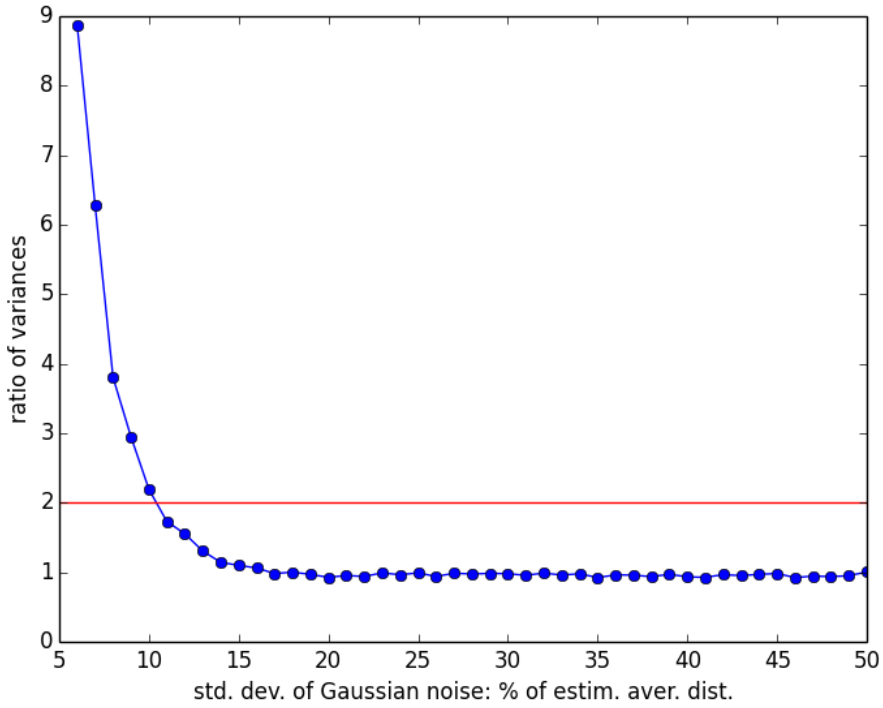

Figure 3.7: The plot shows the quality of discrimination of the crystal types bcc, fcc and hcp for different noise levels of the additive Gaussian noise. This is given in percentage of the estimated average point distance d; 1% d till 5% d are not shown. The parameters been used are $n_{shapes} = 2,\ n_{dir} = 3,\ n_r = 1,\ \{\lambda, q, r\} = \{9, 4, 1.8\}$. The ratio of variances falls below 2 between the noise levels 10% d and 11% d (d is the average point distance). So the method of scaling indices can be used for discrimination between noisy bcc, fcc and hcp till the limit noise level has been reached.

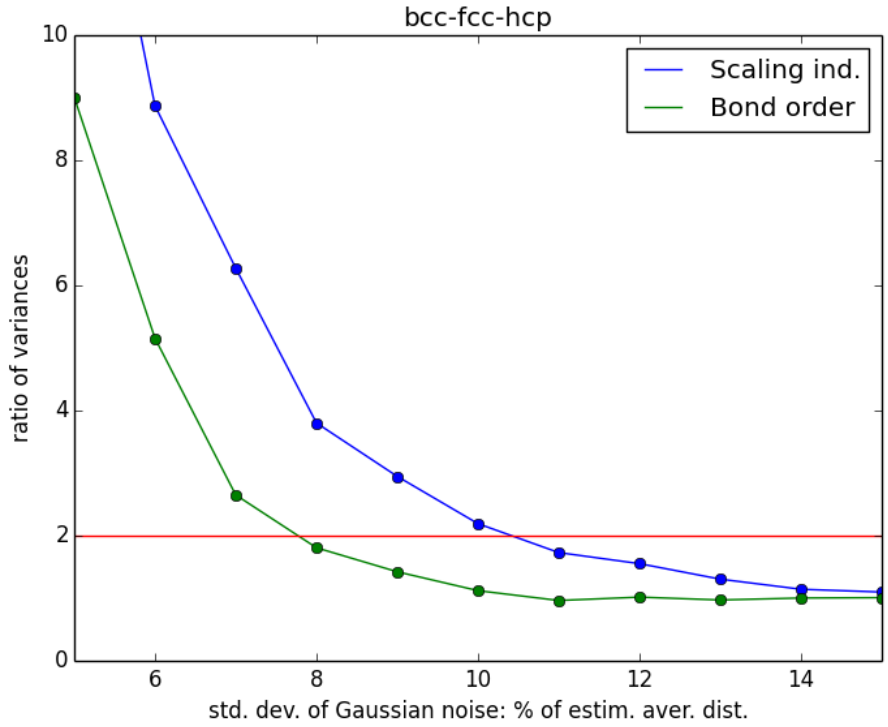

Figure 3.8: This plot shows the quality of discrimination between bcc, fcc and hcp point distributions depending on their noise levels. The scaling indices algorithm (blue) is compared to the bond order parameters algorithm (green). The horizontal red line marks the ration of variances equaling 2 being reached by the scaling indices' graph at about 10% d, by the bond order parameters' graph at about 8% d (d is the average point distance). So the scaling indices algorithm shows a higher stability against noise for the purpose of differing bcc, fcc and hcp types.

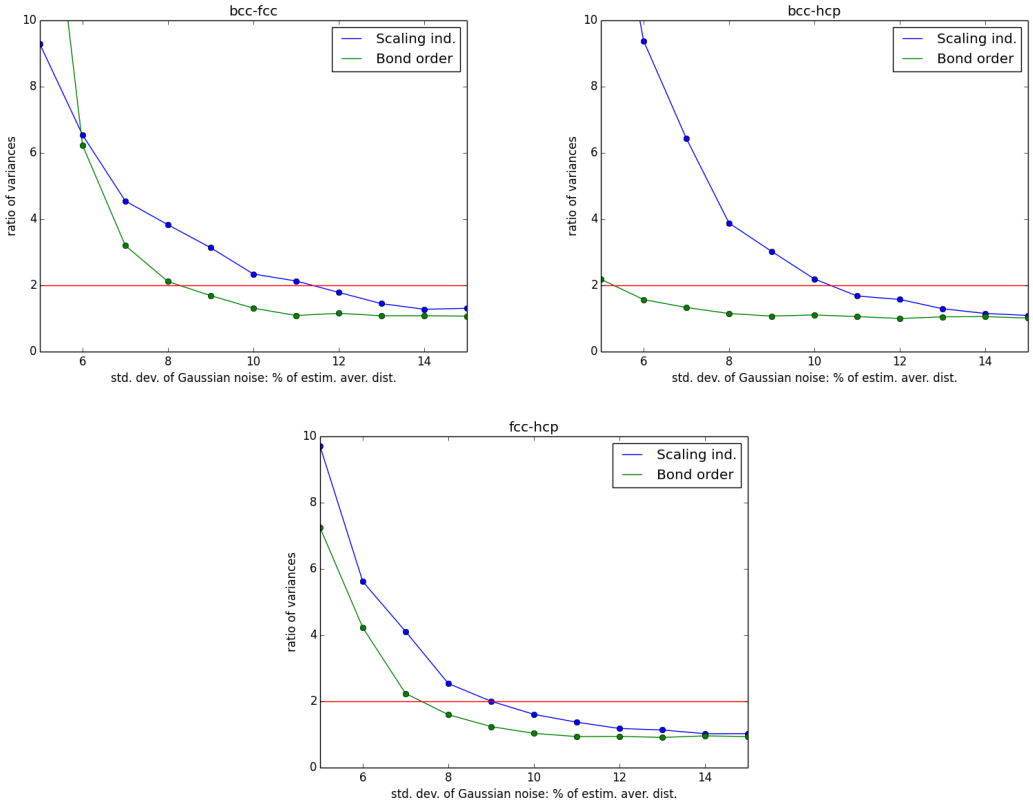

Figure 3.9: This plot shows the quality of discrimination between bcc, fcc and hcp point distrubutions itemized by pairs, and depending on their noise levels. The scaling indices algorithm (blue) is compared to the bond order paremeters algorithm (green). The pairs are bcc - fcc (top left), bcc - hcp (top right) and fcc - hcp (bottom). The scaling indices algorithm always achieves better results regarding the limit noise level for a sufficient discrimination; the limit noise level is between 9% d for fcc – hcp and 11% d for bcc – fcc. Using bond order parameters, the limit noise level already is met between 5% d for bcc – hcp and 8% d for bcc – fcc.

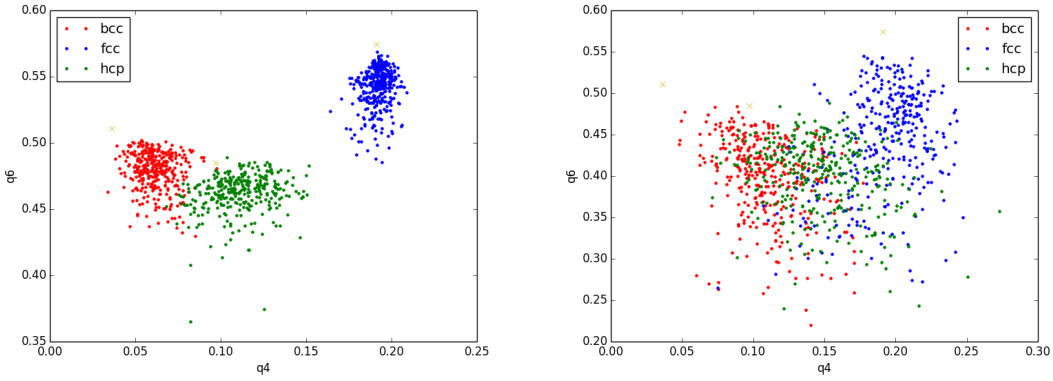

Figure 3.10: This plot visualizes the $Q_4 - Q_6$ parameter space. The results of a bond order parameters analysis of datasets (a) and (b) are described as points in the $Q_4 - Q_6$-space. The red points belong to a bcc distribution, the blue ones to fcc and the green ones to hcp. The left plot underlies a noise level of 5% d (d is the average point distance), the right plot 10% d. In the right plot the different Q_i-distributions clearly overlap, since the limit noise level for a sufficient discriminatio (8% d) is exceeded.

3.6 Determining the parameters λ, q, r for optimal discrimination

The parameters r (characteristic distance), λ (aspect ratio) and q (weighting parameter) need to be determined empirically. Therefore, I made an analysis similar to the one described above for a series of different parameters. The datasets I used were the perfect bcc-, fcc- and hcp-like point distributions with added Gaussian noise with a standard deviation of 5%d (datasets (a)). Finally, I took the values of the parameters, for which the three crystal types could be differed best, i.e. the ratio of variances was the biggest. The parameter r was determined with an accuracy of one decimal place; the parameters λ and q with no decimal place: $\lambda, q \in \mathbb{N}$.

Determining optimal aspect ratio λ

First I determined the optimal aspect ratio λ. Therefore I made a series of analyses for different λ: $\lambda \in \{2, 3, \dots 15\}$. In each analysis the scaling indices of the point distributions were calculated for $n_q = 10$ (apart from this section, q is always fixed, $n_q = 1$), $q \in \{1, 2, \dots 10\}$, and $n_r = 30$, $r \in \{0.1\ d, 0.2\ d, \dots 3.0\ d\}$, for three spatial directions x, y and z, $n_{dir} = 3$, and for both kinds of shapes, $n_{shapes} = 2$. This yielded resulting point clouds in a $n_{par} = 1800$-dimensional α-space. The analyses were repeated 10 times, the distributions differed only regarding the additive noise. The average trend of the ratio of variances from the 10 runs is plotted as solid line in the graph in figure (3.11), the standard deviation is indicated by the error bars. Obviously, the quality of discrimination does not strongly depend on λ, as long as $\lambda > 2$; the best result was obtained for $\lambda = 9$. The

usage of isotropic scaling indices $\lambda = 1$ does not offer any chance for a discrimination, only anisotropic ones are usable.

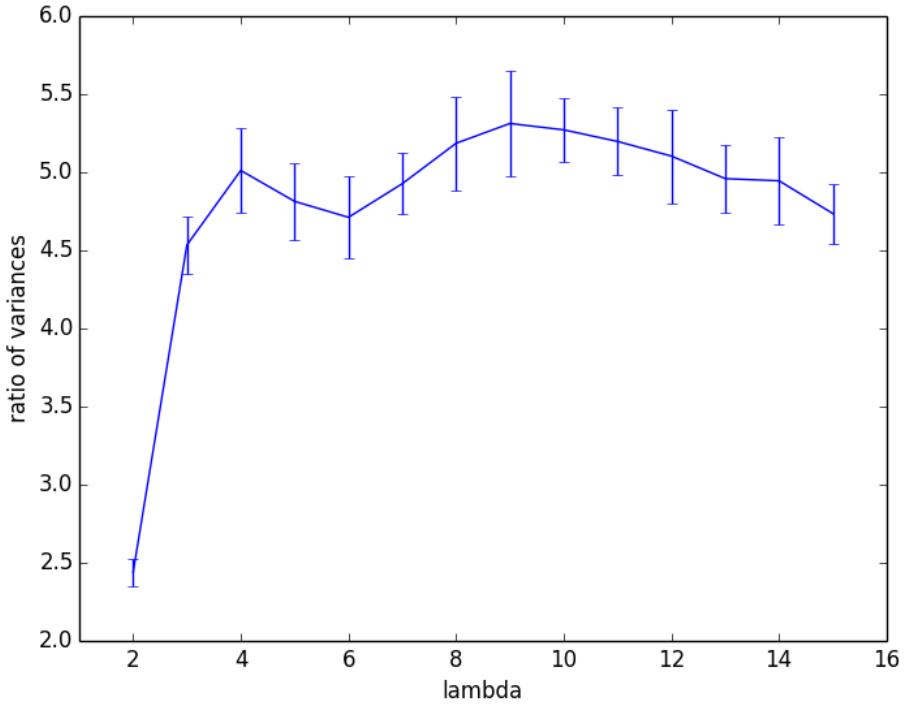

Figure 3.11: Searching for the optimal λ. The plot shows the average quality of discrimination using different λ's (the average is taken from 10 runs using the same point distribution, but with different random noise of equal noise level). The error bar indicates the standard deviation from these 10 runs. Finally, $\lambda = 9$ is chosen.

Determining optimal weighting parameter q

Having fixed the aspect ratio $\lambda = 9$, I made a series of analyses for different weighting parameter $q \in \{1, 2, \ldots 10\}$. In each analysis the scaling indices of the point distributions were calculated for $n_q = 1$, $n_r = 30$, $r \in \{0.1\ d, 0.2\ d, \ldots 3.0\ d\}$, $n_{dir} = 3$, $n_{shapes} = 2$. This leaded to resulting point clouds in a $n_{par} = 180$-dimensional α-space. The analyses were repeated 10 times. The average trend of the ratio of variances from the 10 runs is plotted as solid line in the graph in figure (3.12), the standard deviation is indicated by the error bars. The quality of discrimination is really bad for $q = 1$ and increases till $q = 4$, then decreases again. It seems that scaling indices offer a better chance for discrimination if they are sufficiently local (q controls the slope of the exponential shape of ρ), but still involve distant points beyond the characteristic distance r.

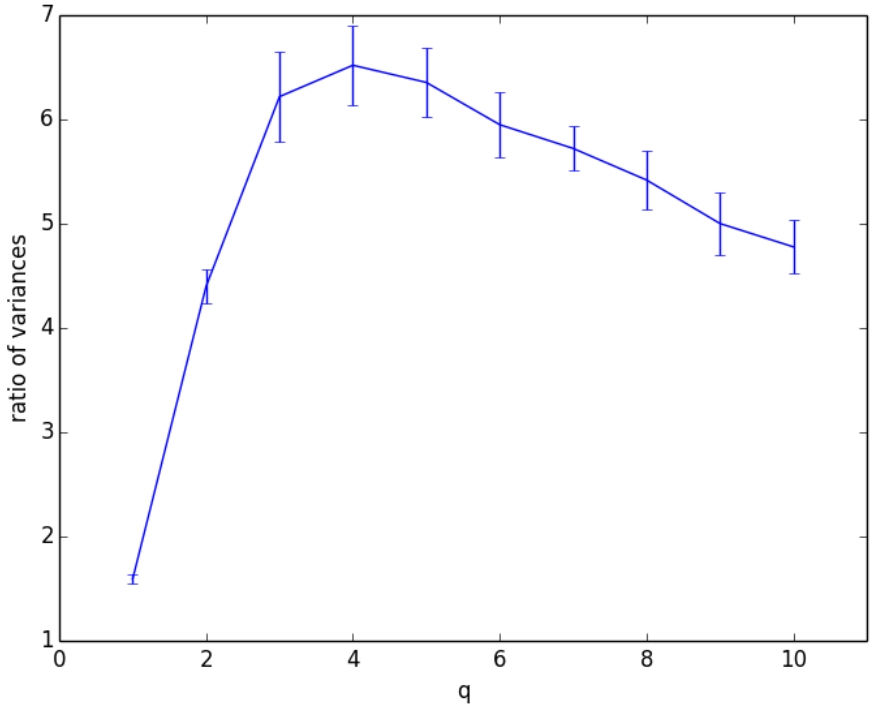

Figure 3.12: Searching for the optimal q. The plot shows the average quality of discrimination using different q (the average is taken from 10 runs using the same point distribution, but with different random noise of equal noise level). The error bar indicates the standard deviation from these 10 runs. Finally, $q = 4$ was chosen.

Determining optimal radius r

Having fixed parameters $\lambda = 9$ and $q = 4$, I repeated the same task to determine the best radius r. I made an analysis using different $r \in \{0.1d, 0.2d, \ldots 3.0d\}$. So $n_r = 1$, $n_{dir} = 3$, $n_{shapes} = 2 \rightarrow n_{par} = 6$. The analyses were repeated 10 times. The average trend of the ratio of variances from the 10 runs is plotted as solid line in the graph in figure (3.13), the standard deviation is indicated by the error bars. Obviously you get peaks at $r = 1.1\ d$ and $r = 1.8\ d$. The different peaks might occur due to differences in the bcc-, fcc-, hcp-elementary cells at according r's. For example, the peak at about $r = 1.1\ d$ might occur due to differences within the first elementary cell. The dip between the peaks might occur due to the same cubic basic structure of bcc and fcc cells with 8 atoms in the corners, etc.

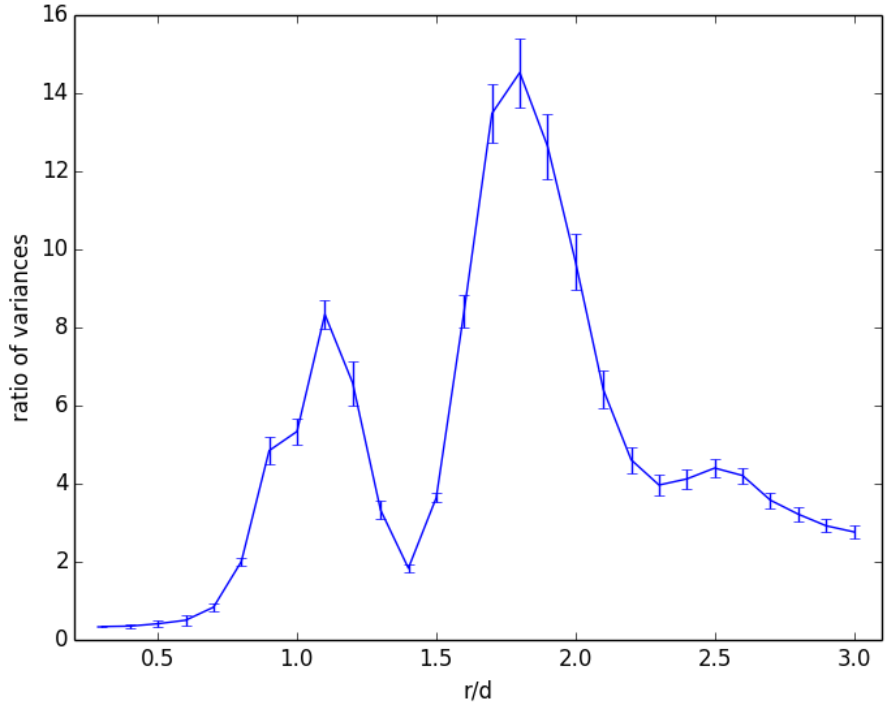

Figure 3.13: Searching for the optimal r. The plot shows the average quality of discrimination using different r (the average is taken from 10 runs using the same point distribution, but with different random noise of equal noise level). The error bar indicates the standard deviation from these 10 runs. Obviously, there are two peaks indicating the shell structure of the crystal types. Finally, for the analyses in this section, $r = 1.8\ d$ was used.

Chapter 4

Scaling indices algorithm for 3D unknown point distributions

The last chapter yielded that scaling indices are usable for the discrimination of crystal types, even a bit better than using bond order parameters. Due to this encouraging result, they shall be used for analyzing real data from experiments and simulations. In the following chapter the requirements and properties of an algorithm used for unknown datasets are described. The algorithm has been employed on three datasets; their description and the results of the analyses are given in the subsequent chapter.

For the experimental datasets the obvious main problem is the complete lack of knowledge, for example whether or not the point distribution contains crystalline structures, the type of crystals, their orientations and possible deformations or the scalings (the absolute value of the distances within a crystalline cell).

A central part of classifying structures is comparing a measured set of scaling indices of point i, $\{\alpha_k\}_i,\ k = 1, \ldots n_{par}$, with a-priori known templates: $\{\alpha_k\}_i \stackrel{?}{\Leftrightarrow} \{\alpha_k\}_{template}$. Since the templates, however, have been measured in artificially created distributions, and the set $\{\alpha_k\}_i$ in an unknown distribution, with both having different spatial scalings, the algorithm needs to scaling free. This means the range $[r_{i,\ ini},\ r_{i,\ fin}]$ for calculating $\{\alpha_k(r)\}_i$ needs to be searched for each single structure i, because it cannot be fixed a-priori sufficiently good in terms of parameters such as the average point distance or the distance of next neighbors. The average point distance can vary in different regions of the point distribution, which can be of different crystal types, more or less populated or non crystalline. The usage of the distance to next neighbors first requires defining next neighbors and second is sensitive on noisy motions of the neighbors.

Furthermore, the second main issue of this chapter is defining "what is crystalline?", i.e. giving a reasonable criterion for taking apart crystalline and non-crystalline structures. For this reason, a criterion of stability is introduced. This ensures that only clusters of similar point structures are considered as being crystalline. Detected crystalline structures then are classified on the basis of a-priori known templates.

Last, the missing rotational invariance of the scaling indices has to be taken into account. The scaling indices of a local structure change when the local structure is rotated. So regarding scaling indices, two structures of the same shape, but of different orientations are considered to be of different types, and therefore need to covered with distinct templates.

4.1 Scaling-free algorithm

In the analyses done so far, $\alpha_i(r)$ was calculated for $n_r = 1$, with fixed $r = 1.5\,d$ or $r = 1.8\,d$. In an unknown data set, which for sure is not as homogeneous as an artificially created one, using the global average point distance d is a bad idea. This can vary strongly in different regions of the point distribution. So my algorithm is set up as follows: The scaling indices in the set $\{\alpha_k\}_i,\ k = 1, \dots n_{par}$ are calculated within a certain range $[r_{i,\ ini}^{sh,\ dir},\ r_{i,\ fin}^{sh,\ dir}]$ which is fixed by certain values for $\alpha_i(r_{i,\ ini}^{sh,\ dir})$ and $\alpha_i(r_{i,\ fin}^{sh,\ dir})$. The abbreviations sh and dir belong to the shapes and directions of $\rho_i(r)$: $sh = cig, dis$ for cigar- or discus-like shape, respectively, and $dir = x, y, z$ for the x, y, z-directions. The number of steps of radius including $r_{i,\ ini}^{sh,\ dir}$ and $r_{i,\ fin}^{sh,\ dir}$ is denoted with n_r. So the algorithm becomes scaling-free, since the values of $r_{i,\ ini}^{sh,\ dir}$ and $r_{i,\ fin}^{sh,\ dir}$ do not matter.

In my algorithm, the cigar-like and the discus-like shape of $\rho_i(r)$ are used ($n_{shapes} = 2$). When calculating the scaling indices with the cigar-like shape, I searched for $r_{i,\ ini}^{cig,\ dir}$ and $r_{i,\ fin}^{cig,\ dir}$, where $\alpha(r_{i,\ ini}^{cig,\ dir}) \approx 0.3\ \forall i, dir$ and $\alpha(r_{i,\ fin}^{cig,\ dir}) \approx 2.5\ \forall i, dir$. These values of α are attained only once in the $r - \alpha(r)$-curve, as you can see in figure (4.1). Therefore they can be found quite easily by iteration. The outcome of this algorithm is shown in the first half of figure (4.3). The curves shown in this plot belong to perfect bcc-, fcc- and hcp-crystals.

When calculating the scaling indices with the discus-like shape, the problem occurs that the related $r - \alpha(r)$-curve has its characteristic features around $\alpha = 3$ (fig 4.2), why an adequate and stable $r_{i,\ fin}^{dis,\ dir}$ cannot be found easily by iteration. Therefore I made use of the already calculated values of $r_{i,\ ini}^{cig,\ dir}$ and $r_{i,\ fin}^{cig,\ dir}$ belonging to the cigar-like shape and used their means for the discus-like shape:

$$r_{i,\ ini}^{dis,\ x} = r_{i,\ ini}^{dis,\ y} = r_{i,\ ini}^{dis,\ z} = mean(\{r_{i,\ ini}^{cig,\ x}, r_{i,\ ini}^{cig,\ y}, r_{i,\ ini}^{cig,\ z}\}),\ \forall i \tag{4.1}$$

$$r_{i,\ fin}^{dis,\ x} = r_{i,\ fin}^{dis,\ y} = r_{i,\ fin}^{dis,\ z} = mean(\{r_{i,\ fin}^{cig,\ x}, r_{i,\ fin}^{cig,\ y}, r_{i,\ fin}^{cig,\ z}\}),\ \forall i \tag{4.2}$$

So for the cigar-like shape one gets a set $\{\alpha_k\}_i,\ k \in \{1, \dots 300\}$, since only the main spatial directions x, y, z are considered, $n_{dir} = 3$, and 100 steps of r are used, $n_r = 100$, including $r_{i,\ ini}^{cig,\ dir}$ and $r_{i,\ fin}^{cig,\ dir}$. Analogously, for the discus-like shape one gets a set $\{\alpha_k\}_i,\ k \in \{301, \dots 600\}$. The overall structure of the resulting array is shown in figure (4.4). A template set visualized in figure 4.3 shows the outcome of the definitions for $r_{i,\ ini}^{cig,\ dir}$, $r_{i,\ fin}^{cig,\ dir}$, $r_{i,\ ini}^{dis,\ dir}$, $r_{i,\ fin}^{dis,\ dir}$.

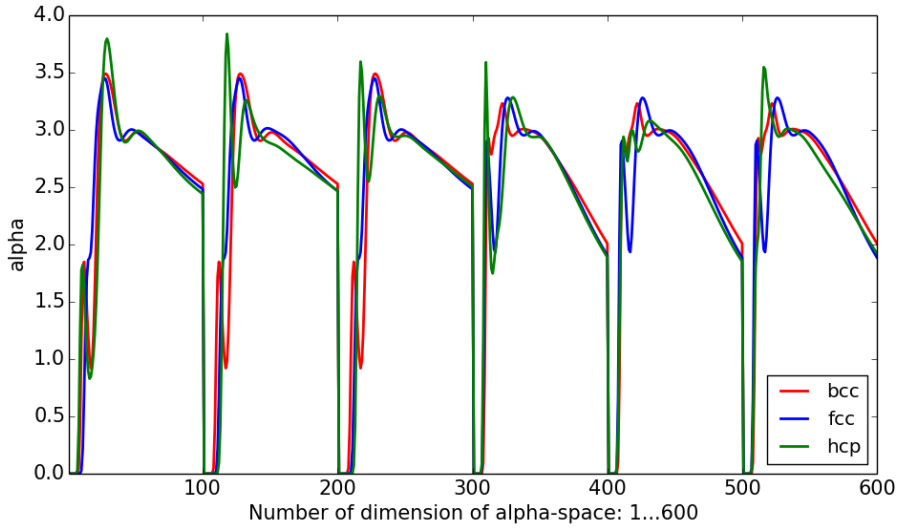

Figure 4.1: This plot shall motivate the definition of $r_{i,\ ini}^{cig,\ dir}$ and $r_{i,\ fin}^{cig,\ dir}$ with $\alpha(r_{i,\ ini}^{cig,\ dir}) \approx 0.3\ \forall i, dir$ and $\alpha(r_{i,\ fin}^{cig,\ dir}) \approx 2.5\ \forall i, dir$. Aim is to find a region of $r - \alpha(r)$ which offers a good chance of discrimination between the crystal types.
The plot shows the set $\{\alpha_k\}_i,\ k = 1, \dots n_{par}$. The array structure (also sketched in fig 4.4) is as follows: The dimensions $[1, 300]$ belong to the cigar-like shape, $[301, 600]$ to the discus-like shape. Each 100 dimensions belong to different directions x, y, z. The range of radius r for each 100 dimensions ($[1, 100],\ [101, 200], \dots$) is $[0.1\ d,\ 10\ d]$, d is the average point distance. The plot shows the result only for a certain orientation of the point distribution. For the cigar-like shape (dimensions $[1, 300]$), α shows clear differences between the crystal types in the range $\alpha \in [0, 3]$, the values $\alpha = 0.3$ and $\alpha = 2.5$ are attained only once, why I chose them for defining $r_{i,\ ini}^{cig,\ dir}$ and $r_{i,\ fin}^{cig,\ dir}$. For the discus-like shape (dimensions $[301, 600]$), α does not show differences between the crystal types for $\alpha < 3$ (also see 4.2). Therefore, an analogue definition cannot be made.

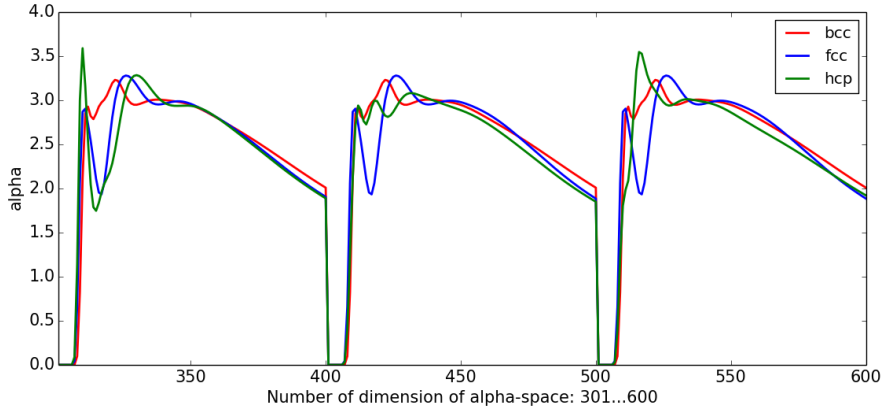

Figure 4.2: This plot shows the curves $r - \alpha(r)$ calculated using the discus-like shape. Dimensions in the ranges $[301, 400]$, $[401, 500]$ and $[501, 600]$ belong to different directions x, y, z (see fig 4.4 about array structure.) The interesting structures of the graphs start only for $\alpha(r) > 3$, so an approach as for the cigar-like shape is not possible (searching a range $[r_{i,\ ini}^{cig,\ dir}, r_{i,\ fin}^{cig,\ dir}]$ with $\alpha(r_{i,\ ini}^{cig,\ dir}), \alpha(r_{i,\ fin}^{cig,\ dir}) \in [0, 3]$ that covers the interesting features). The plot shows the result for a certain orientation of the point distribution.

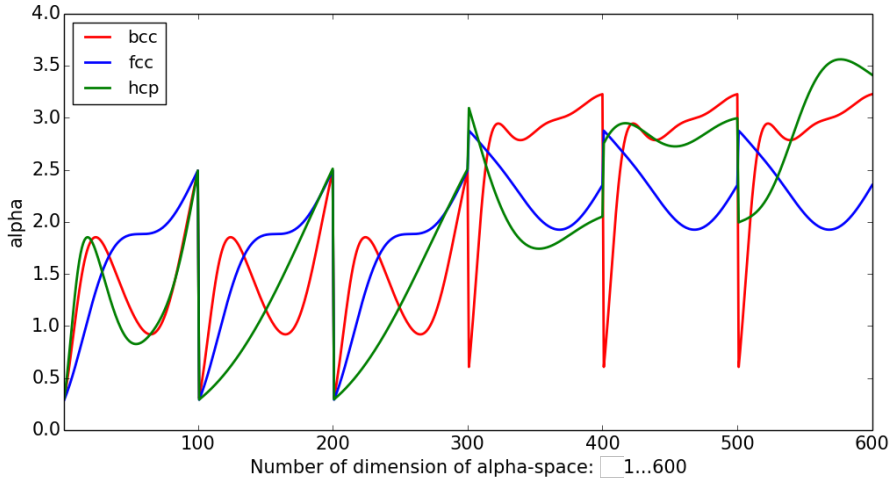

Figure 4.3: This plot shows the outcome of the definition of $r_{i,\ ini}^{sh,\ dir}$ and $r_{i,\ fin}^{sh,\ dir}$. The curves belong to the cigar-like (dimensions $[1, 300]$) and discus-like (dimensions $[301, 600]$) shapes. The ranges of the radius for each 100 entries are the related $[r_{i,\ ini}^{cig,\ dir},\ r_{i,\ fin}^{cig,\ dir}]$ or $[r_{i,\ ini}^{dis,\ dir},\ r_{i,\ fin}^{dis,\ dir}]$, respectively. The plot shows the result for a certain orientation of the point distribution.

However, this method of determining $r^{dis,\ dir}_{i,\ ini}$ and $r^{dis,\ dir}_{i,\ fin}$ is not optimal. Furthermore, there are no structures detected in the range $\alpha \in [0, 3]$, which might prove the discus-like shape a disappointing choice. More about this in the chapter about results 5.5.

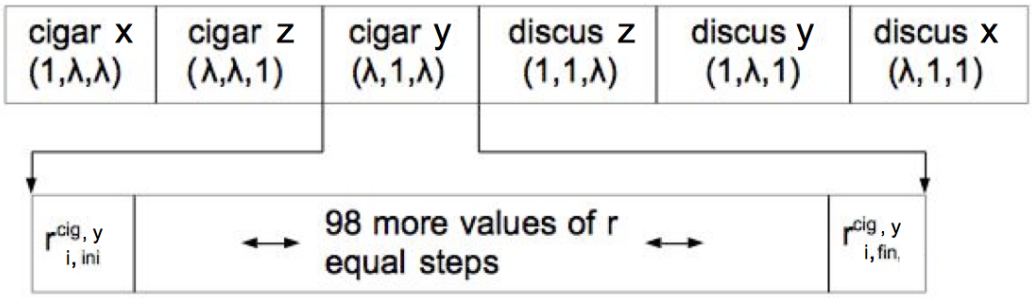

Figure 4.4: This image shows the structure of an array containing the set of scaling indices of a point in a 3D point distribution. Since the α-space has $n_{par} = n_{shapes} \cdot n_{dir} \cdot n_r = 2 \cdot 3 \cdot 100 = 600$ dimensions, the array has 600 entries, the first 300 for the cigar-like shape, the second 300 for the discus-like one. The sets of eigenvalues of the distance measure $\{\lambda_p\}$ leading to the according shape of ρ are stated in each block (always with aspect ratio $\lambda = 9$). Each block contains 100 scaling indices $\alpha_k(r)$ with $r \in [r^{sh,\ dir}_{i,ini},\ r^{sh,\ dir}_{i,fin}]$. The abbreviation $sh = cig, dis$ refers to the cigar-like or discus-like shape of $\rho_i(r)$; $dir = x, y, z$ refers to the direction of $\rho_i(r)$. E.g. cig, y means $\rho_i(r)$ has the shape of a cigar aligned along y-direction; dis, y means $\rho_i(r)$ has the shape of a discus with rotational axis in y-direction.

4.2 Defining "crystalline"

For each point i of a point distribution a set of scaling indices $\{\alpha_k\}_i$, $k = 1 \ldots n_{par}$ was calculated; these sets can be taken as single points in the n_{par}-dimensional α-space. The main and most doubtful issue of searching for crystalline structures is to find a reasonable criterion whether or not a structure can be taken as crystalline.

One simple way of doing this is searching the most similar template $t(i)$ for all points i in the distribution, so the distance in α-space, $\Delta^{\alpha}_{i,\ t(i)}$, needs to be minimal. The notation $t(i)$ shall indicate that template t is the most similar for point i, i.e. the template t with $\min_t^{templates} \Delta^{\alpha}_{i,t}$. For taking apart "crystalline" and "non-crystalline" structures, a threshold could be introduced, so that $\Delta^{\alpha}_{i,\ t(i)} \overset{!}{<} threshold$. This, however, is not reasonable. Also in random point distributions local structures can be more or less similar to one of the templates. Furthermore, a distribution may contain large clusters of homogeneous shape, whose $\Delta^{\alpha}_{i,\ t(i)}$ all exceed the threshold, though. For clarifying this, several histograms were calculated. A histogram of $\{\Delta^{\alpha}_{i,\ t(i)}\}_i$ of all points $i = 1, \ldots N$ should show large differences between point distributions with many crystalline structures and point distributions with a little number of crystalline structures or random distributions. In the PK-3-Plus dataset consisting of eleven time frames, where crystallization sets in,

however, the histograms of different scans do not show these large differences (figure 4.5); not at all between each other, and not significantly compared with the histogram of a random distribution. For comparison: The number of detected crystalline structures using the bond order parameters method for the PK-3-Plus data set differs strongly between Scan 1 and Scan 9 (figure 5.7). This yields that a threshold for differing "crystalline" and "non-crystalline" cannot be set.

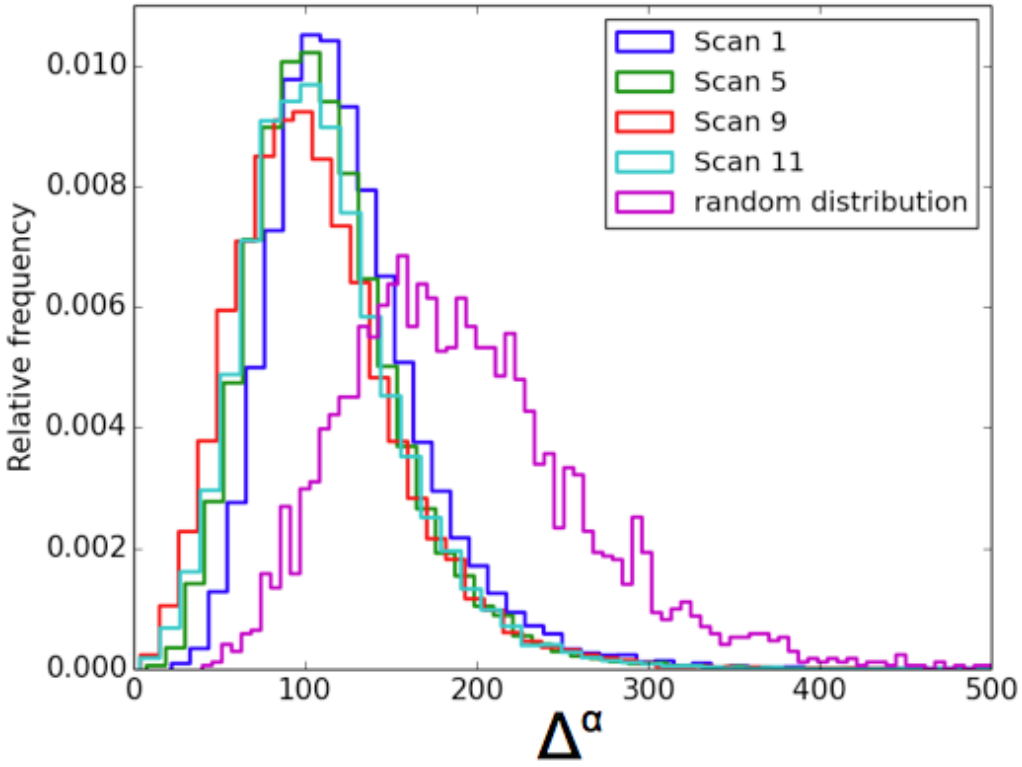

Figure 4.5: These histograms show the relative frequencies of $\{\Delta^{\alpha}_{i,\ t(i)}\}_i,\ i = 1, \ldots N$ in four scans of the PK-3-Plus dataset and in a random point distribution, where index i runs through all points of each point distribution, and index $t(i)$ means the corresponding most similar template to point structure i: $\min_t^{templates} \Delta^{\alpha}_{i,t}$.

So a better criterion for differing "crystalline" and "non-crystalline" needs to be found. A good criterion should require the stability of a local structure in order to call it crystalline. A stable local structure needs to fulfill the crystalline characteristic of translational order. The criterion, however, shall keep the local character of the algorithm. I used following criterion: I defined a "crystalline" structure as a local structure with a similar neighborhood. "Similar" is meant in terms of a little distance in α-space, Δ^{α}. As a reminder, the distance in the parameter space between points i and j is:

$$\Delta^{\alpha}_{ij} = \sum_{k=1}^{n_{par}} (\alpha_{i,k} - \alpha_{j,k})^2 \tag{4.3}$$

(Whether or not to take the square root is not important; in the algorithm it is not done due to time efficiency. n_{par} denotes the dimensionality of the α-space, which is for analyses of 3D point distributions 600 and for 2D point distributions 120.)

The criterion is fulfilled if:

$$\frac{1}{N_{neighbors}} \sum_{j(i)=1}^{N_{neighbors}} \Delta^{\alpha}_{i,j(i)} < threshold \tag{4.4}$$

"Neighbors" does not necessarily mean only next neighbors in the first shell. So there are two parameters to fix: the *threshold* of similarity ("how similar shall the neighbors be?") and $N_{neighbors}$ ("how many neighbors shall be compared with?"). To fix the values the following criterion is used: In the artificially created points distributions which are originally perfect crystalline with some additive Gaussian noise, as many as possible crystalline structures shall be found for small noise (5% d, with d being the average point distance); for large noise (15% d) as many as possible local structures shall be allocated to "non crystalline":

$$\begin{aligned} \text{low noise point distribution} &\xrightarrow{!} \text{crystalline} \\ \text{high noise point distribution} &\xrightarrow{!} \text{non-crystalline} \end{aligned}$$

The value $5\%d$ was chosen without special reason. I considered it to be small enough, also due to the results in the chapter 3, where the 5% d point distribution demonstrated a good discrimination between the crystal types. The value 15% d was chosen first because of the result in figure (3.7), where the ratio of variances reaches 1 at about 15% d; second because according to Lindemann's criterion [42] a crystalline structure melts when the strength of the particle fluctuations δ around their mean position, $\sqrt{<\delta^2>}$ (which is the standard deviation of the Gaussian noise in the artificially created point distributions), reaches a value of more or less 15% of the average point distance.

To keep this criterion the parameter $N_{neighbors}$ has very little effect. The much larger effect has the similarity *threshold*. Figure (4.6) shows the fractions of detected crystalline structures in low-noise point distributions of bcc, fcc or hcp type (solid lines) and the fractions of non-crystalline structures in high-noise point distributions of bcc, fcc or hcp type (dashed lines). Both fractions shall be maximal simultaneously; values around $threshold \approx 80$ fulfil this criterion. $N_{neighbors}$ was chosen to be 18; the number of first plus second shell next neighbors is 18 for fcc and hcp and 14 for bcc, so a value of 18 is still quite local.

Finally, the parameters have been fixed: $threshold = 80$, $N_{neighbors} = 18$. In this chapter these parameters are not modified. In the section 6.2 it is played around with them, since the effects can be recognized easily for 2D point distributions.

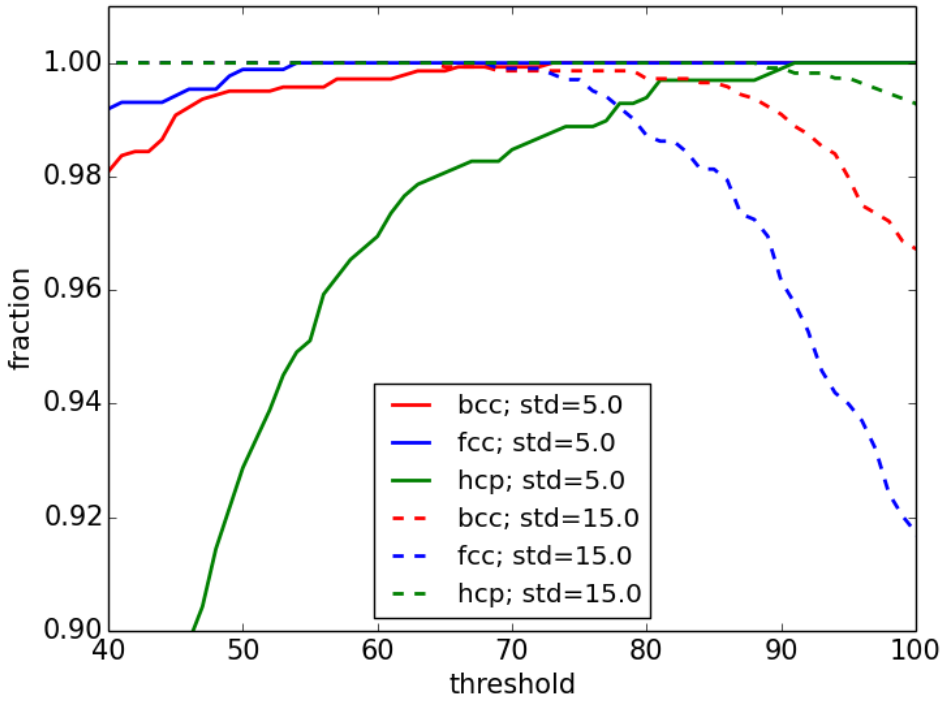

Figure 4.6: This plot shows the fraction of detected crystalline structures in low noise (noise level 5% average point distance) point distributions (solid lines) and the fractions of non allocated structures in high noise (noise level 15% average point distance) point distributions (dashed lines), depending on the similarity *threshold*. Both fractions shall be maximal simultaneously which is complied for *threshold* ≈ 80.

4.3 Classifying crystalline structures

The detected crystalline structures shall be further classified. Therefore, these local structures are compared to a-priori known templates of certain crystal types, namely of bcc, fcc and hcp type. They are allocated to the template which has minimal distance in α-space. So for each point i and each template t the distance $\Delta^{\alpha}_{i,t}$ is calculated; then point i is allocated to template $t(i)$ with the minimal distance $\min_{t}^{templates}\{\Delta^{\alpha}_{i,t}\}_t$.

A basic set of templates contains the sets of scaling indices for perfect point distributions of bcc, fcc and hcp type, without discrimination of angular orientations of the point distribution more precise than the 3D main directions x, y, z.

Due to the fact of not being rotationally invariant, a set of anisotropic scaling indices, which were measured only along the 3D main axes, changes when the point distribution (e.g. a bcc crystal) is rotated (fig 4.7). This leads to following problems: when rotating e.g. a bcc crystal in many ways, which was originally orientated $\theta = \phi = 0$ in spherical coordinates (arbitrary reference orientation), a large part of resulting orientations differ strongly among themselves regarding Δ^{α}. So if these orientations are not saved as templates a-priori, they may not be identified as being of bcc type: e.g. in a set of 24 different orientations of a originally $\theta = \phi = 0$ oriented bcc crystal, only 4 orientations are identified as being of bcc type, 16 as fcc and 4 as hcp.

So obviously the set of templates has to be extended in order to contain several different spatial orientations of all crystal types. Saved as templates were the following orientations: for the bcc and fcc crystals all combinations of $\theta, \phi \in \{0, \frac{\pi}{4}, \frac{\pi}{2}\}$. For the hcp crystal all combinations of $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\}$ and $\phi \in \{-\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}\}$.

In order to avoid degeneracies among the templates, i.e. two or more different crystalline structures or orientations generate the same (or at least a similar) set of scaling indices, it is actually necessary to measure the scaling indices along many more spatial directions than only the 3D main axes (different rotation matrices in the distance measure, see figure 4.8). This was not done in this thesis for analyses of 3D point distributions, only for analyses of 2D point distributions. The usage of the discus-like shape for the calculation of the scaling indices, however, lifts many degeneracies quite well. The degeneracies due to the intrinsic symmetry of bcc, fcc and hcp are of course left.

The similarity between the templates, **with** usage of the discus-like shape, is shown in figure (4.10); for comparison: figure (4.9) shows the similarity between the templates **without** using the discus-like shape. The values in the tables show $\Delta^{\alpha}_{t,\,t'}$ expressed as percentage of the similarity threshold, i.e. 100 (and above) meaning not similar at all, and 0 totally equal. For reasons of clarity, values beyond 100 are set to 100. Indices $t,\ t'$ refer to templates: $t, t' \in \{\text{all templates}\}$. Comparing the two tables, obviously when also using the discus-like shape (4.10) overall the templates get less similar (in terms of Δ^{α}): much more often the value 100 is reached and small values below 50 are not attained any more.

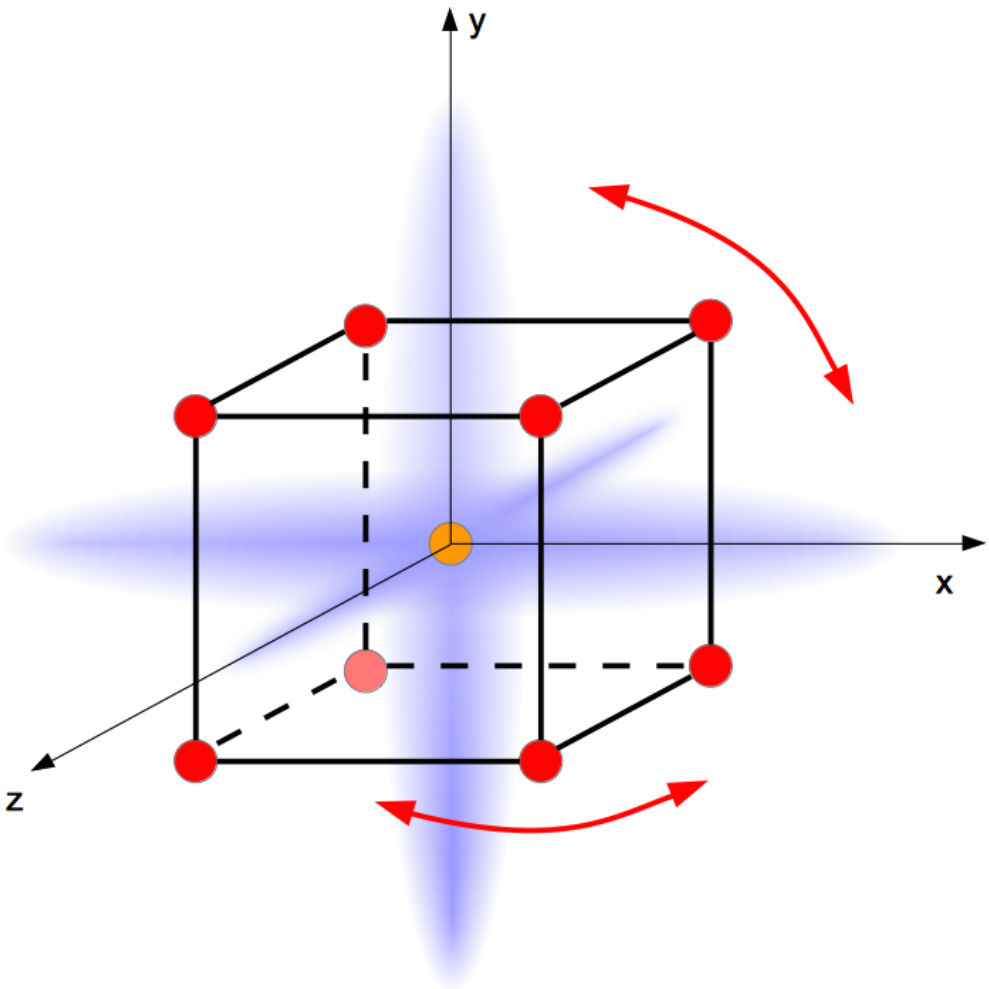

Figure 4.7: This plot shall illustrate a measuring of scaling indices along three main directions x, y, z (blue ellipses). These directions are fixed. The point distribution (orange and red points) is rotated along θ- and ϕ-direction around the orange middle point, indicated by the red arrows. This leads to following problem: The measurements give very different results for different orientations of the point distribution, so these different orientations need to be saved as templates for a correct allocation to the crystal types. Arising degeneracies need to be lifted by also using different rotations matrices in the distance measure $||\cdot||_A$ (see introduction 1.4), as indicated in figure (4.8), or by using the discus-like shape of $\rho_i(r)$.

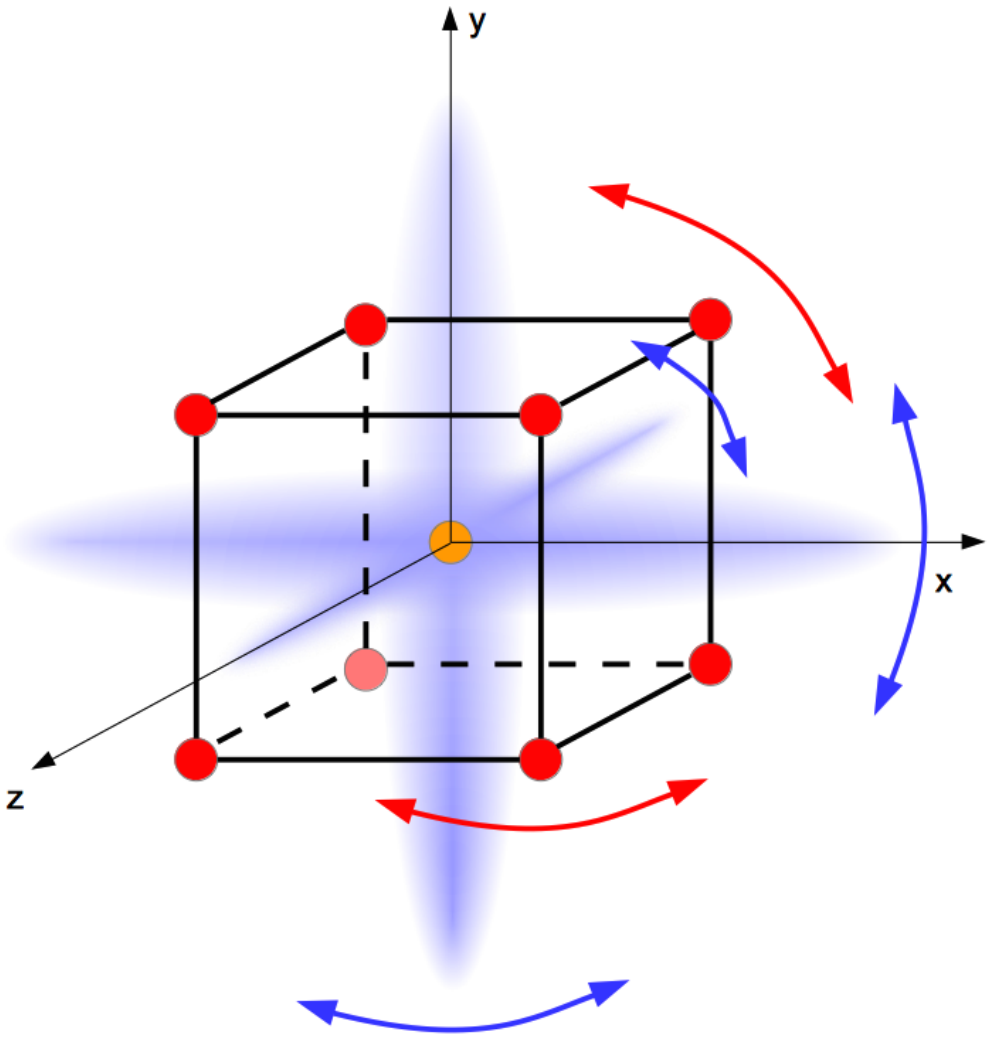

Figure 4.8: This plot sketches the usage of different rotation matrices in the distance measure $||\cdot||_A$ (see introduction 1.4), indicated with the blue arrows. Rotations of the point distribution are indicated by the red arrows.

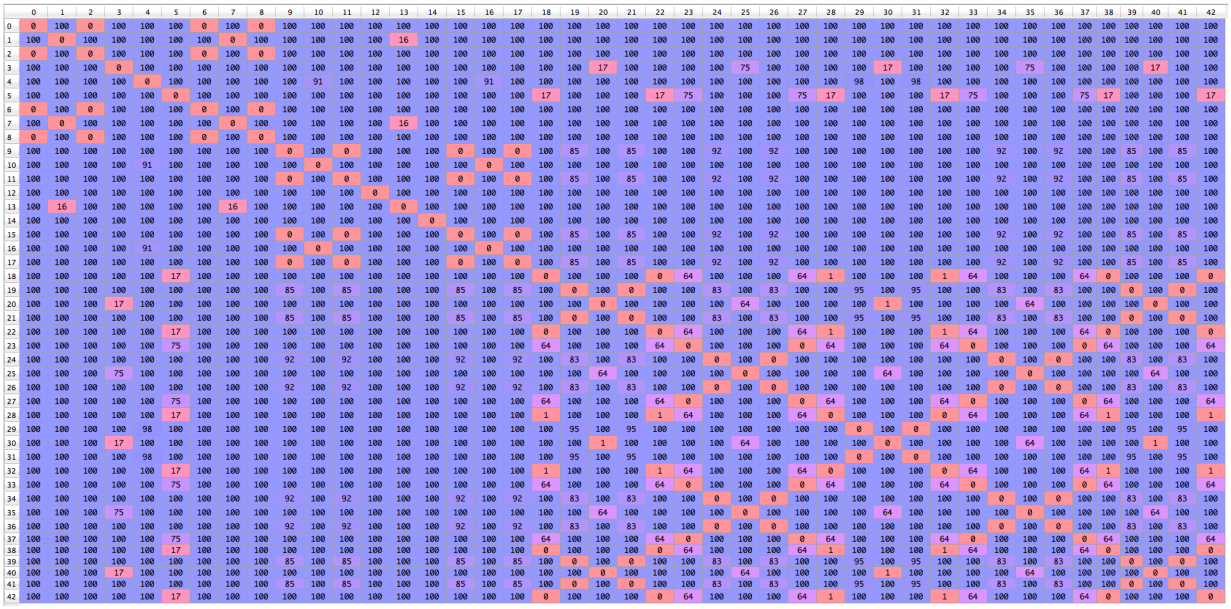

Figure 4.9: This table shows the similarity of the templates calculated **without** using the discus-like shape. The rows and columns refer to the templates of bcc (nine templates: 0-8), fcc (nine templates: 9-17) and hcp (25 templates: 18-42). The values refer to $\Delta^{\alpha}_{t,\,t'}$ ($t,\ t' \in$ {all templates}) and are given in the percentage of the threshold of similarity. So 0 means that templates t and t' are equal, 100 means not similar at all. Values larger than 100 are set to 100.

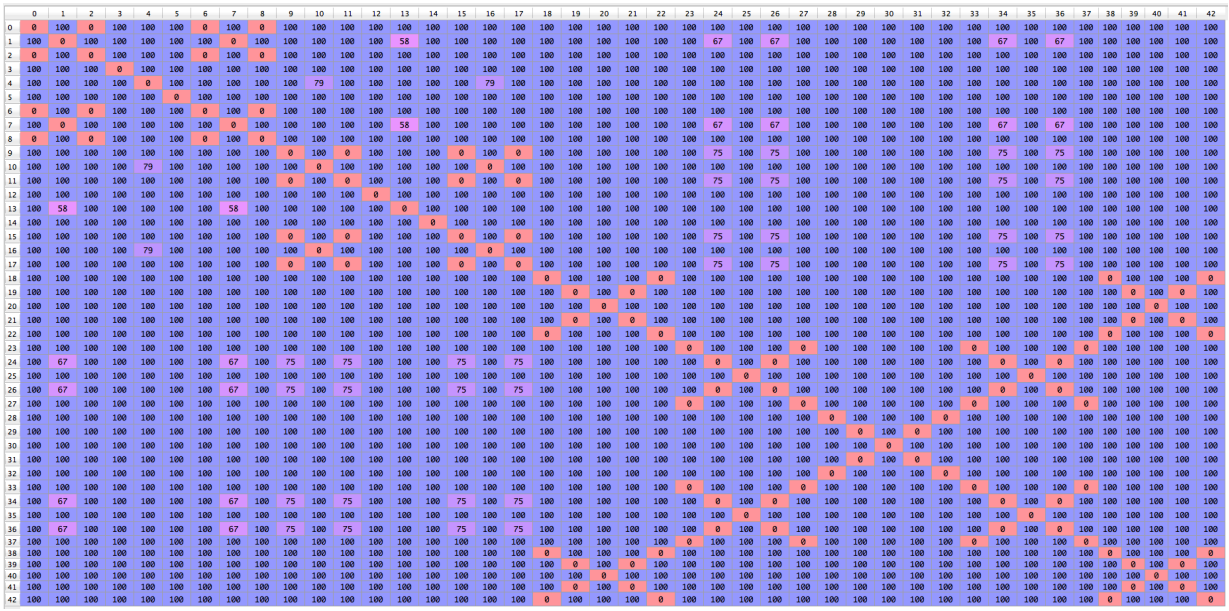

Figure 4.10: This table shows the similarity of the templates calculated **with** using both shapes of $\rho_i(r)$. The rows and columns refer to the templates of bcc (nine templates: 0-8), fcc (nine templates: 9-17) and hcp (25 templates: 18-42). The values refer to $\Delta^{\alpha}_{t,\,t'}$ ($t,\ t' \in$ {all templates}) and are given in the percentage of the threshold of similarity. So 0 means that templates t and t' are equal, 100 means not similar at all. Values larger than 100 are set to 100. Compared with the previous table (4.9), here the values of Δ^{α} more often reach 100, so the templates get less similar if using both shapes of $\rho_i(r)$ instead of only one. The degeneracies due to the symmetric structures of the crystal types are left.

Chapter 5

Results for 3D unknown point distributions

The algorithm described in the last chapter has been employed on three datasets. It has classified all local structures in the distributions as certain crystal types or as "non-allocated", respectively. Discussions of the results are made within the next sections for the different datasets and afterwards at the end of the chapter.

5.1 Description of analyzed datasets

The datasets used in this chapter are as follows:

- Artificially created 3D point distributions of either bcc, fcc or hcp type with certain noise levels
- Dataset obtained from PK-3-Plus experiments containing eleven 3D point distributions [43] [44]
- 3D point distribution from PKE Nefedov experiments [45] [46]

The artificially created point distributions are used as basis for setting up the algorithm. They are created from the same perfect point distributions as described in section 3.1, only the additive Gaussian noise here has a standard deviation that reaches from 1% up to 15% of the average point distance.

The second dataset, which was obtained in PK-3-Plus experiments, consists of eleven frames shot at different times with each about 20k points. The dataset was created and analyzed using bond order parameters by Peter Huber, Ludwig-Maximilians-Universität München, [44]. The PK-3-Plus [43] chamber (illustration in 5.1), which was used in microgravity on board of the International Space Station, exists of a cuboid glass cuvette and contains two aluminium rf-electrodes, where gas is brought inbetween and been ionized. Reservoirs of argon and neon gas and a system for gas in- and outlet enable the usage of

pure argon or neon gas or a mixture. With six particle dispensers in total, micro particles are interspersed into the plasma. The micro particles got charged due to the interaction with the plasma and arrange in crystalline shapes due to their strong electrostatic potential and their confinement. The experiment resulting into this dataset, was performed on the 20th mission on 9th January 2013, Run 3. Silica particles of diameter $1.55\mu m$ and argon gas were used. The gas pressure was $\approx 10Pa$, the peak-to-peak voltage of the electrodes $\approx 81.5V$. The dataset contains the 3D spatial positions of these particles (figure 5.2), which are recorded by a laser-camera-system. The lasers have a wavelength of $686nm$ and an adjustable power of $40mW$. Four cameras with PAL standard, a resolution of 768×576 pixels and a rate of $25Hz$ or $50Hz$ allow a 3D view of the complex plasma as well as the observation of the glow characteristics of the plasma. Much more detailed information about the experimental setup of PK-3-Plus and the data collection are to be found in the stated reference, as well as a tabular comparison of PK-3-Plus and PKE-Nefedov experiments.

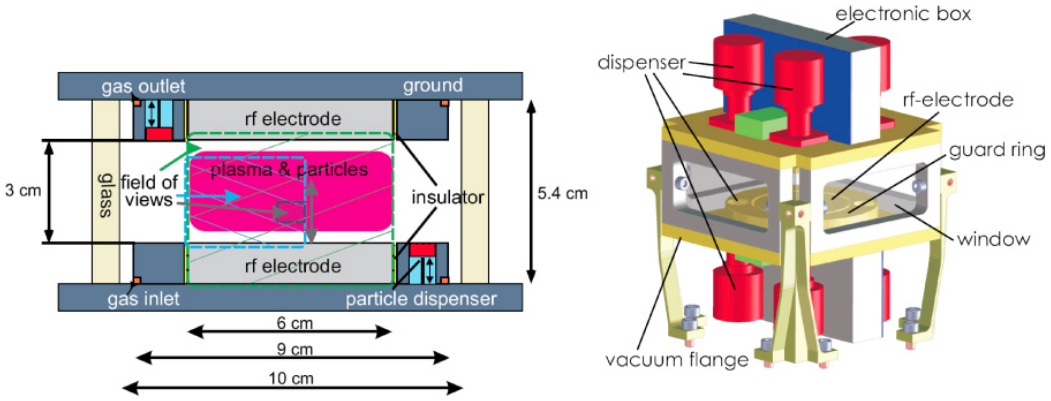

Figure 5.1: The PK-3-Plus chamber was run in micro-gravity on board of the ISS for producing and watching complex plasmas. On the left a cross-sectional, on the right a perspective view are presented.

The third dataset (created and analyzed using modified bond order parameters by Christopher Dietz, Justus-Liebig-Universität Giessen) as well is experimentally obtained. It contains a 3D distribution of about 30k particle positions of a complex plasma created in a plasma chamber of type PKE Nefedov [45]. For the detailed description of the experimental setup see the master thesis of C. Dietz [46]. The parameters which were used for this dataset are: diameter of the melamine formaldehyde particles: $1.31\mu m$; gas pressure: $28.5Pa$; power: $0.16W$. The point distribution is shown in figure (5.3).

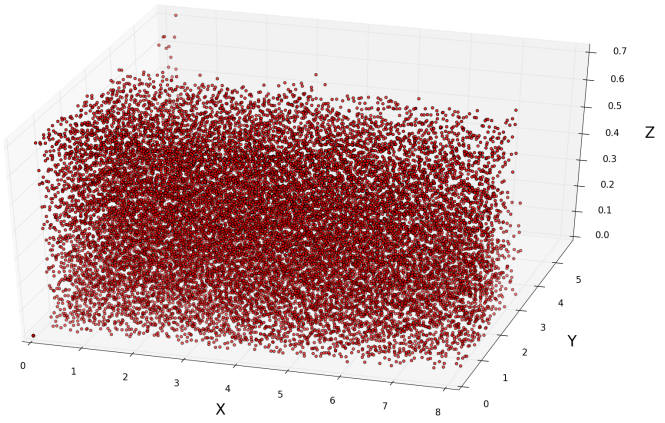

Figure 5.2: All scans of the PK-3-Plus dataset contain 3D particle positions of about 20k particles. This plot illustrates the first point distribution of the set (scan 1). The other scans look similar globally. The border region, which is two average point distances thick, was not analyzed.

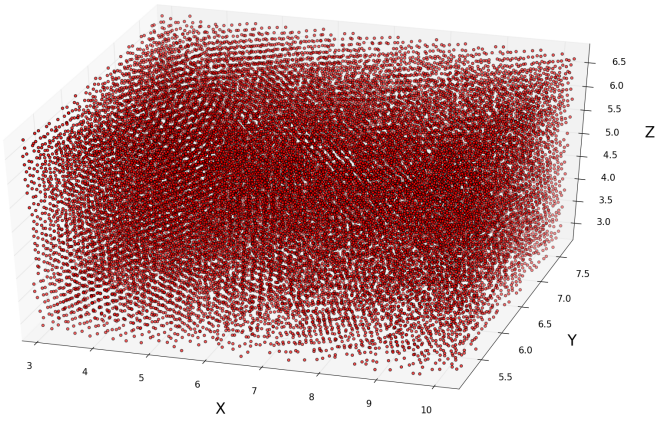

Figure 5.3: This plot shows the particle positions of the data obtained from the PKE-Nefedov experiment by C. Dietz. It contains about 30k points; the border region was not analyzed.

5.2 Results for dataset 1: Artificially created point distributions

The artificially created point distributions were used as basis for setting up the algorithm. So the results are not very surprising; they are shown in figure (5.4) for point distributions of originally perfect bcc type, figure (5.5) for point distributions of originally perfect fcc type and figure (5.6) for point distributions of originally perfect hcp type. The border region with a thickness of two average point distances was not analyzed.

For low noise, in bcc, fcc and hcp distributions all points were allocated to bcc, fcc or hcp type, respectively. For higher noise, more and more local structures start being allocated to different crystal types; this is strongest for the fcc point distributions starting at a noise level of about 5% d (with d being the average point distance); at a noise level of about 10% d, about a quarter of the point distribution is allocated to hcp type. With the noise level reaching 15% d no more crystalline structures are found, as it was defined a priori (see section 4.2).

The relatively large number of detected hcp crystals in the fcc series can either be due to real "transitions" in the local structures from fcc type to hcp type for higher noise; or the reason is just a bad measuring of the local structures. Problems could occur due to the usage of the discus-like shape and due to a too small number of directions the scaling indices are measured along (as described in sections 4.3 and 5.5).

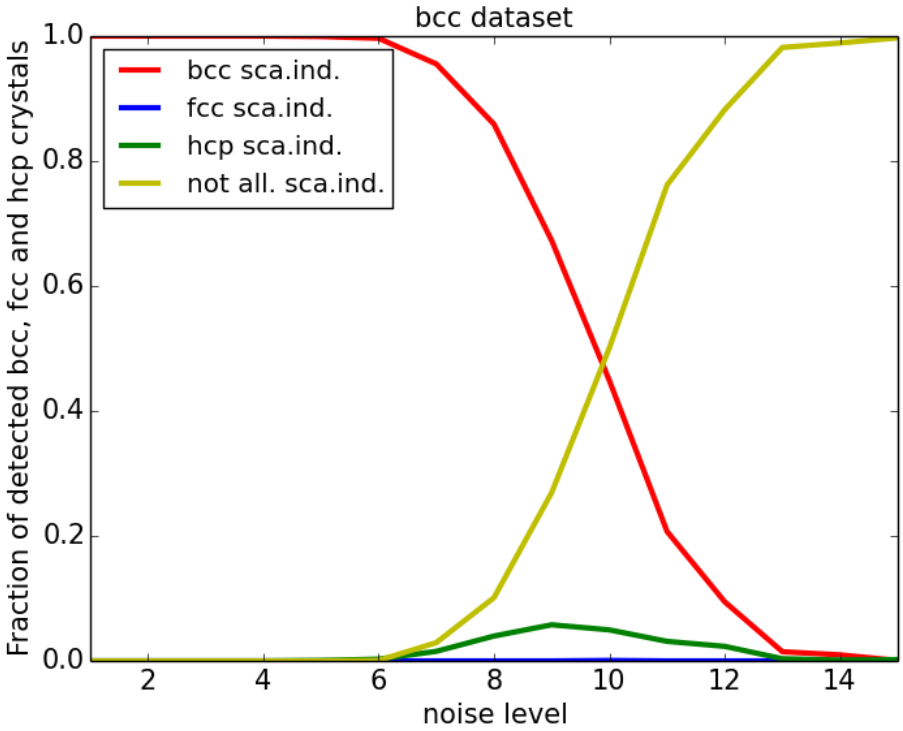

Figure 5.4: The fractions of detected bcc (red), fcc (blue) and hcp (green) crystals in originally perfect and pure point distributions of **bcc** type. To these points Gaussian noise was added with a standard deviation from 1% up to 15% of the average point distance, respectively. The number of non-allocated structures is drawn yellow. The local structures in the point distribution are predominantly allocated to bcc type, for all noise levels. Between 6% d and 13% d a fraction of max. about 0.05 is allocated to hcp type.

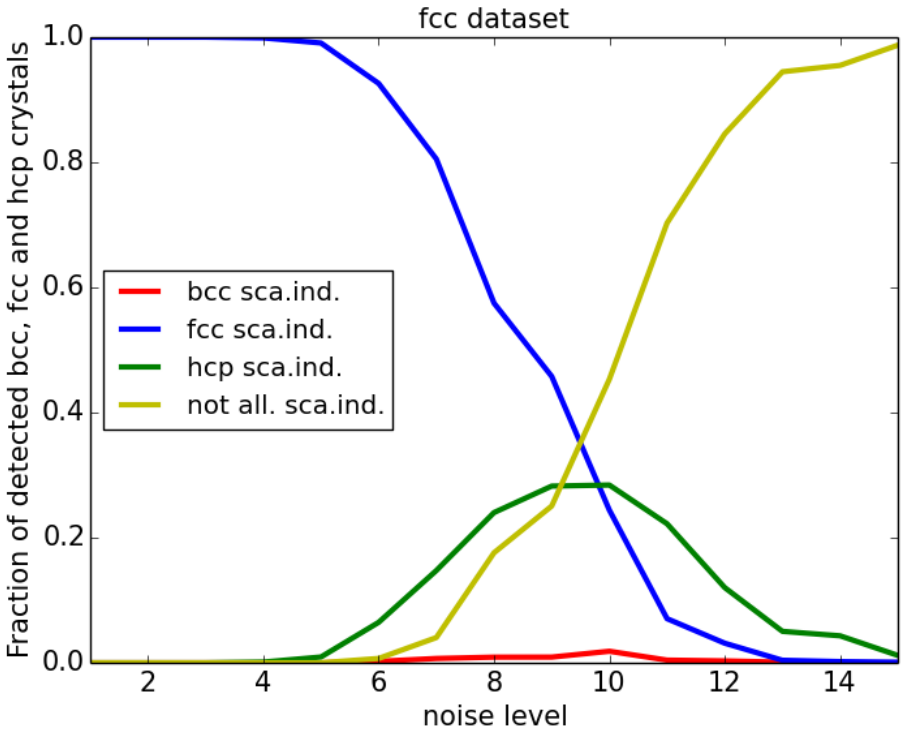

Figure 5.5: The fractions of detected bcc (red), fcc (blue) and hcp (green) crystals in originally perfect and pure point distributions of **fcc** type. To these points Gaussian noise was added with a standard deviation from 1% up to 15% of the average point distance, respectively. The number of non-allocated structures is drawn yellow. Starting at a noise level of about 5% d (with d being the average point distance), local structures are also allocated to hcp type. The maximum fraction of about one quarter is reached at a noise level of about 10% d. With the noise level reaching 15% d no more crystalline structures are found, as it was defined a priori (see section 4.2

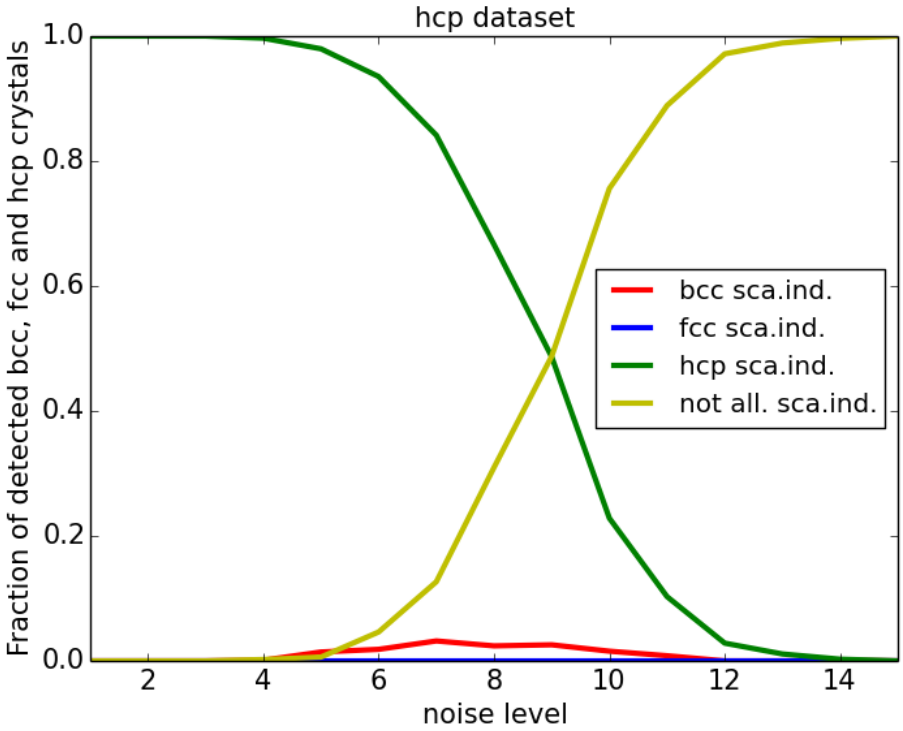

Figure 5.6: The fractions of detected bcc (red), fcc (blue) and hcp (green) crystals in originally perfect and pure point distributions of **hcp** type. To these points Gaussian noise was added with a standard deviation from 1% up to 15% of the average point distance, respectively. The number of non-allocated structures is drawn yellow. Only very tiny fraction of the local structures is allocated differently than to hcp, for all noise levels.

5.3 Results for dataset 2: from PK-3-Plus experiments

The point distributions of the PK-3-Plus dataset were analyzed with the scaling indices method and the bond order parameters method. The plot in figure (5.7) shows the numbers of detected crystalline structures found with the scaling indices algorithm (solid lines) and found using the bond order parameters method (dashed lines) in 11 time frames. With bond order parameters, however, no bcc type was searched for; the border region with a thickness of two average point distances was not analyzed. The analyses using bond order parameters were first performed with an algorithm by Peter Huber ([44]) and repeated with own code in Python. Values for bond order parameters as well as for thresholds were adopted from Peter Huber. The algorithm from Peter Huber additionally contains a stability criterion for differing non-crystalline, crystalline bulk- and crystalline surface-particles, identified dependent on the number of next-neighbor-crystals, which are just counted after classifying all particles dependent on the values of the bond order parameters. This criterion was not adopted; the results in my thesis are only on basis of the values of the bond order parameters.

The result for Scan 2 were illustrated as well as 3D colored plots of the particle positions (bcc crystals red, fcc blue and hcp green). Figure 5.8 shows the crystals (and only the crystals) detected using the scaling indices method, figure 5.9 using bond order parameters. The appendix contains these kind of plots for all Scans (see 7.5 and following).

Common results in both methods are:

- the strongly increasing number of hcp crystals between Scan 1 and Scan 6;
- the number of hcp crystals is non-zero and equal in Scan 1;
- the overall trend of the number of hcp crystals is very similar;
- the number of fcc crystals in Scan 1 equals 0;
- the trend of the number of non-allocated structures: it sharply declines at the beginning, has a minimum at about Scan 8 and rises again in the end.

Differences are:

- the overall gap regarding the numbers of detected hcp and fcc: the method using bond order parameters detected about 1000 more hcp crystals and 500 more fcc crystals overall in all scans;
- the bends in the trend of the number of fcc crystals are different, e.g. in Scan 3 it is the other way around.
- (the number of bcc crystals)

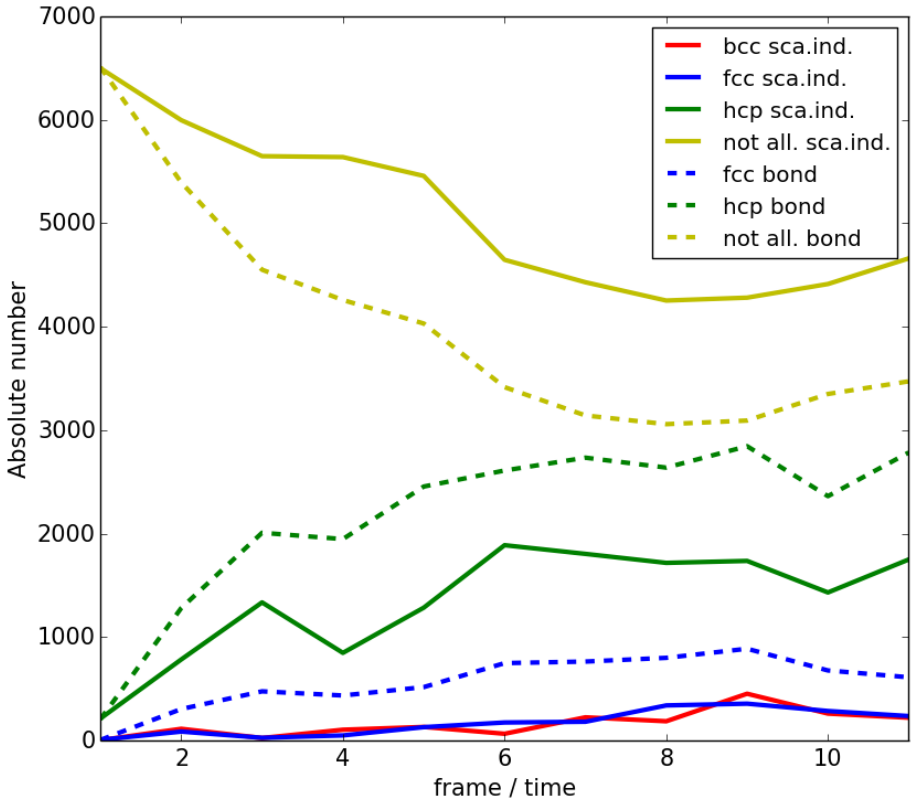

Figure 5.7: This graph shows the resulting numbers of detected crystals of bcc (red), fcc (blue) and hcp (green) type, as well as the number of non allocated structures, obtained by the scaling indices method (solid lines) and with bond order parameters (dashed lines). For a comparison of the graphs, see list and text.

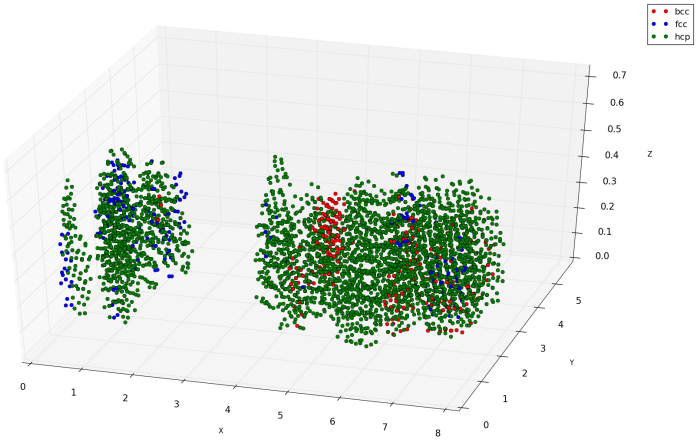

Figure 5.8: Here the point distribution of Scan 2 of the PK-3-Plus dataset is illustrated. Only crystalline structures detected using the **scaling indices method** are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

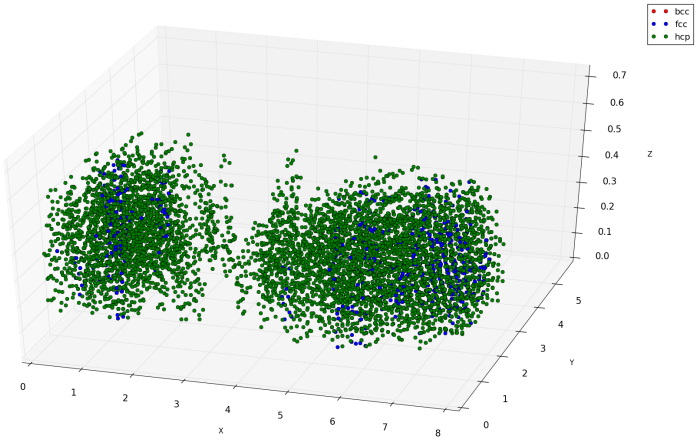

Figure 5.9: Here the point distribution of Scan 2 of the PK-3-Plus dataset is illustrated. Only crystalline structures detected using the **bond order parameters method** are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

The result obtained with the bond order parameters method overall shows a larger number of crystals. In chapter 3, however, it was told that the scaling indices method was able to discriminate between the crystal types till a higher noise level than the method of bond order parameters. This actually should lead to a larger number of detected crystals when using the scaling indices method. Those point distributions used in that chapter were known a priori, though. For unknown data sets – such as the dataset from PK-3-Plus – the stability criterion was introduced. This criterion ensures that no crystalline structures are found any more in a high noise point distribution, even if there are perfect single cells. This lowers the number of detected crystalline structures in point distributions compared with a fully local algorithm, because only a local structure being surrounded by a similar neighborhood is counted as crystalline. "Unstable" local structures are sorted out.

The results of my scaling indices algorithm are not itemized by the categories bulk-particles and surface-particles, because I have not set up such an additional allocation routine. This could for example be done equally to the algorithm from Peter Huber. A surface-particle then just is identified by defining a number of necessary non-crystalline particles in the next neighborhood of a crystalline particle. Even not having made this discrimination, the behavior over time of the number of detected fcc and hcp particles, does not cause a change in the physical interpretation. Important facts, namely the number of fcc and hcp structures in Scan 1, and the rising number of fcc and hcp particles over time, are equal. So, according to Huber [44], the crystallization of the distribution seems to start at hcp-like nuclei, which grow over time. Furthermore, Huber assumes the hcp type changes into fcc type, during the complex plasma looses energy (cools down), and so fcc is less energetic (the ground state). Whether or not there really is a transition from hcp to fcc, however, cannot be told with my algorithm, because I did not record the behavior over time of the allocated crystal type for the single points (this cannot be done just by comparing the results of different Scans. First the trajectories of the particles need to be tracked, but due to the small number of frames and too large time step, the movements of the particles are too choppy in this dataset). Furthermore, the implication of the behavior over time of the bcc number is an interesting issue. In Scan 1, the bcc number equals 0, equally to the fcc number, and then rises, too. The bcc type has a smaller packing density than the close-packed types fcc and hcp; this actually suggests a transition from bcc to the other types, when cooling down. An interesting question is, whether the bcc type crystallizes in the surroundings of the hcp nuclei, and then changes into the other types, or crystallizes seperately. This can be resolved as well when analyzing a dataset with much more and denser time frames, giving access to the tracking of particle trajectories and time recording of the allocation types.

The matches and differences of both methods regarding the allocation of the local structures to "crystalline" and "non allocated", with and without distinction of crystal types, are presented in the following two tables (for the results of Scan 9 of the PK-3-Plus dataset). The first one shows, without distinction of crystal type and expressed as percentage of the total number of points, that about three quarters of the points are allocated to equal categories "crystalline" and "not allocated" by both methods. One

quarter of the points is allocated differently. This one quarter mainly exists of points which are "non allocated" by the scaling indices method, whilst allocated to "hcp" by the bond order parameters method.

Values as percentage of total number of points		Scaling indices: crystalline	Scaling indices: non allocated
Bond order	crystalline	33.3	21.4
	non allocated	4.0	41.3

Values as percentage of total number of points		Scaling indices: bcc	fcc	hcp	non allocated
Bond order	fcc	1.6	2.2	7.5	1.7
	hcp	4.6	2.5	14.9	19.6
	non allocated	0.5	0.5	3.0	41.3

An obvious problem is the dissatisfying mismatch in the allocations to fcc and hcp crystal type. A clear trend cannot be identified; fcc and hcp from one method are allocated to both fcc and hcp by the other method. This maybe comes from the problematic usage of the discus-like shape or - what is tantamount - the too less number of different spatial directions that the scaling indices are calculated along. This also explains the already mentioned discrepancy between the good results using the scaling indices in chapter 3 and here. The artificially created point distributions in chapter 3 were known a-priori and have a certain orientation. For this orientation the scaling indices can differ between the crystal types very well as described there. This good discrimination, however, can only be obtained, if this "well-working spatial direction / dimension" is handled by the scaling indices, i.e. the scaling indices are measured also along this direction. This is not ensured in the algorithm, if the number of directions is too small.

5.4 Results for dataset 3: from PKE-Nefedov experiments

Excursion: Introduction to the algorithm of C. Dietz

The algorithm of Christopher Dietz (group "Plasma and Astronautics Physics", Prof. Markus Thoma, Justus-Liebig-Universität Giessen), as well uses the Bond order parameters (the description of the algorithm was taken from Dietz' presentation at the DSDP Workshop Kiel 2015 [47] – also see [46]; the results were taken from private communication). Before calculating Q_4, Q_6, by means of Q_2 solid and liquid local structures are differed. Then, in addition to the standard algorithm, the space containing the point distribution is subdivided in its Voronoi cells (see e.g. [48]). For the calculation of the Bond order parameters Q_4, Q_6 of any point i, the spherical harmonics Y_{lm} of the next neighbors of point i are weighted with the corresponding facet areas $A(f),\ f \in F(i)$, of the Voronoi polyhedron $F(i)$ of point i (so f refers to a facet of F or the corresponding next neighbor of i, respectively).

So the formula for Q_l alters into:

$$Q_l(i) = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left(\sum_{f \in F(i)} \frac{A(f)}{A} Y_{lm}(\theta_f, \phi_f) \right)^2 \right)^{\frac{1}{2}} \tag{5.1}$$

end of excursion

The following table shows the matches and differences of both methods regarding the allocation of the local structures to "crystalline" and "non allocated", expressed as percentage of the total number.

Values as percentage of total number of points		Scaling indices: crystalline	Scaling indices: non allocated
Bond order Dietz	crystalline	66.9	33.1
	non allocated	0	0

The next table itemizes by crystal type:

Values as percentage of total number of points		Scaling indices bcc	fcc	hcp	non allocated
Bond order Dietz	bcc	1.4	0	4.3	4.1
	fcc	4.6	0	49.7	24.4
	hcp	0.5	0	6.3	4.5
	non allocated	0	0	0	0

The remarkable difference in the numbers of fcc and hcp type detected by the scaling indices algorithm and the algorithm of Dietz needs to be clarified. There are some possible reasons for different allocations.

About a third of the analyzed structures was classified as non-crystalline by the scaling indices algorithm, while the algorithm of Dietz allocated most of this third to fcc type. So either the fcc clusters in the point distribution are are much more inhomogeneous regarding spatial order (and so not "similar" in the sense of section 4.2), or the fcc type is just the garbage dump of the Bond order parameters algorithm. The clusters of fcc type, however, per se are not more likely to be inhomogeneous than clusters of bcc or hcp type.

Furthermore, the controversial allocations can come from the fact already shown in figure (5.5) that a relatively large fraction of originally fcc type structures are classified as hcp by the scaling indices algorithm if the noise level is sufficiently high. The possible reasons for this are explained in the next section 5.5.

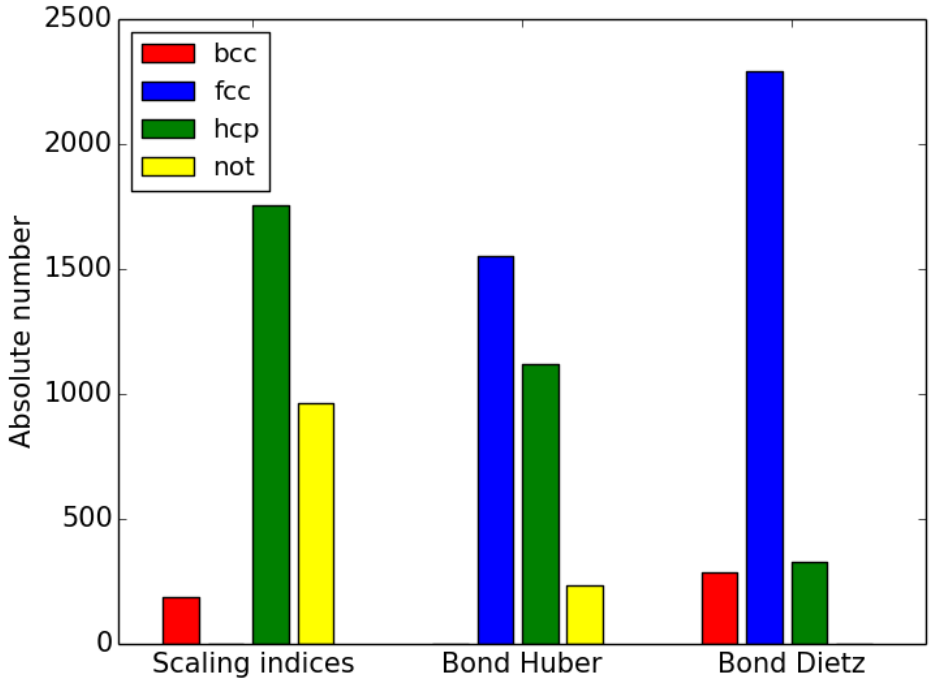

Figure 5.10: Comparison between the results of different methods for the dataset Dietz. The bars show the absolute numbers of detected types of crystals and non-crystalline structures in the analyzed region of the point distribution. The algorithm Bond Huber did not search for bcc crystals. The difference in the allocations regarding fcc and hcp are obvious.

5.5 About problems and solutions

Compelled by the the starkly different results regarding the allocation to fcc- or hcp-type between the bond order parameters and the scaling indices algorithms, I made a few more tests of my algorithm. First, I will show a Principal Component analysis of the α-space containing the results of Scan 9 of the PK-3-Plus dataset, which was one of the most crystalline distributions. The PCA shall show the influence of the different shapes and radii used for the calculation of the set $\{\alpha_{k,i}(r)\},\ k = 1 \ldots n_{par},\ i = 1 \ldots N$ on the discrimination of the crystal types. Second, I will run the algorithm on a set of artificially created point distributions of almost perfect fcc type (very low noise), differing only in the orientations.

Problems with the discus-like shape

I made a Principal Component Analysis of the α-space of Scan 9 in the PK-3-Plus dataset (figure 5.11). I used Scan 9 because it contains the most crystalline structures. The graph shows the entries of the $n \times m = 600 \times 3$ PCA projection matrix V which is composed of the eigenvectors $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ for eigenvalues $\mu_1 > \mu_2 > \mu_3$. They represent the strength of share of each dimension k in the n_{par}-dimensional α-space in its first, second and third Principal Components $\tilde{\alpha}_{h=1},\ \tilde{\alpha}_{h=2},\ \tilde{\alpha}_{h=3}$, respectively. The entries of $\vec{v}_h$, $v_{h,k}$, $h = 1, 2, 3,\ k \in [1,\ 300]$ belong to the cigar-like shape; $k \in [301, 600]$ belong to the discus-like shape.

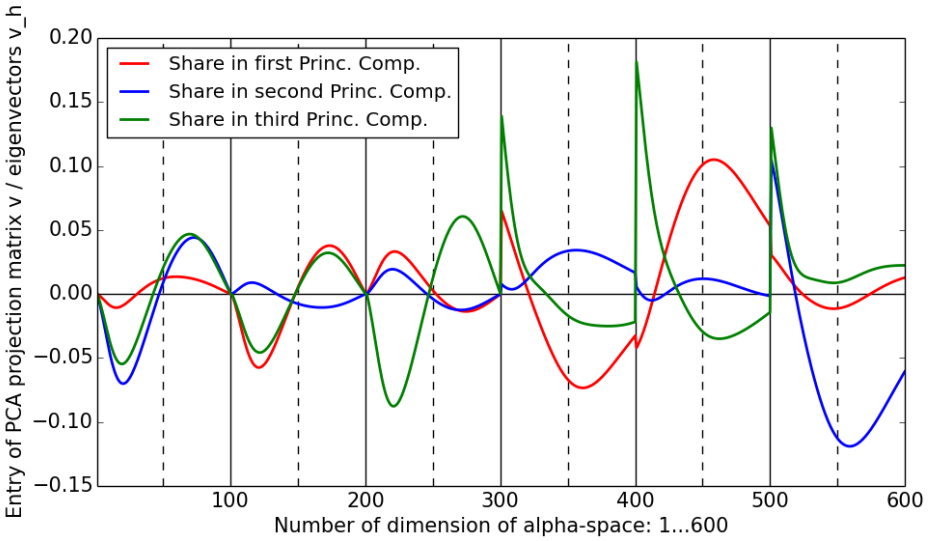

Figure 5.11: This graph shows the entries of the PCA eigenvectors $\vec{v}_1$ (red), $\vec{v}_2$ (blue) and $\vec{v}_3$ (green) belonging to the first, second and third principal components, respectively. They give an insight into how strongly the single dimensions k of the α-space contribute to the discrimination of different crystal types in Scan 9 of the PK-3-Plus dataset.
The contribution of the discus-like shape ($k \in [301,\ 600]$) is very strong, compared with the cigar-like shape ($k \in [1,\ 300]$). This suggests that the discus-like shape causes many noisy fluctuations in the α-space, and does not have a large potential for the discrimination purpose between the crystal types. Especially the large values at e.g. $k = 301, 401, 501$ mean that the values of $\alpha(r_{i,\ ini}^{dis,\ dir})$ differ strongly between all points $i = 1, \ldots N$ of the point distribution. So one can reasonably expect that the sets of scaling indices $\{\alpha_k\}_i,\ k \in [301,\ 600]$ obtained with the discus-like shape are not comparable to template sets.

The values of $v_{h,k}$ at $k = 1, 100, 101, 200, 201, 300,\ h = 1, 2, 3$ are approximately 0, because the algorithm fixes $\alpha_i(r)$, when measuring the scaling indices with the cigar-like shape, at $\alpha(r_{i,\ ini}^{cig,\ dir}) \approx 0.3\ \forall i, dir$ and $\alpha(r_{i,\ fin}^{cig,\ dir}) \approx 2.5\ \forall i, dir$. Hence, there is no variance in the α-space along these dimensions. The minima and maxima in the range $k \in [1, 300]$ belong to values of average values of α of approximately 1.0 and approximately 1.6-1.8, respectively. At $k = 50, 150, 250$ (the dashed lines), the average value of $\langle\alpha\rangle_i \approx 1.4$. This

affiliation can be looked up in figure (5.12), which shows the mean and standard deviation of α over all points in Scan9, itemized by dimension.

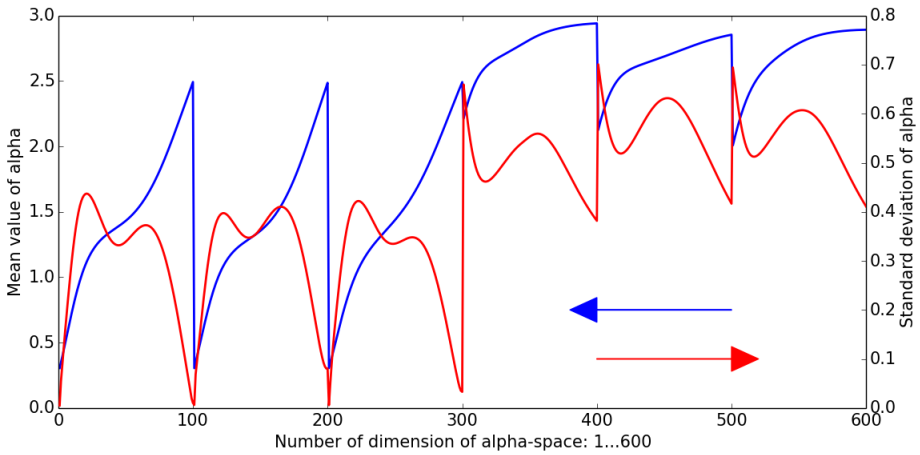

Figure 5.12: This graph shows mean, $\langle\alpha\rangle_i$, and standard deviation of α itemized by dimension, based on all points in Scan 9 of the PK-3-Plus dataset. The values of the mean shall give an insight into what ranges of α are processed in the algorithm; the values of the standard deviation are linked to the graphs of the PCA (figure 5.11): dimensions with big variance contribute much to the discrimination of crystal types.

Looking at the graphs, obviously the discus-like shape has a quite large share in the first three Principal Components. This either might be due to their suppositious much higher potential of discrimination between the crystal types, or due to much more fluctuations caused by them, compared to the cigar-like shape. The second reason might be more likely. The graphs of $r - \alpha(r)$ show characteristic differences between the crystal types only after α has reached 3 (see figure 4.2), i.e. the discus-like shape of $\rho(r)$ gathers a huge number of points at small scales (small r) leading to a sharp rise of $\alpha(r)$, too sharp for a discrimination between the crystal types on small scales r. Figure (5.12, second half) shows basically the same: the graphs of $r - \alpha(r)$ (blue) run between $\langle\alpha\rangle_i \approx 2$ and $\langle\alpha\rangle_i \approx 3$, so the small scales (small r) are not handled by the algorithm when using discus-like shape.
Especially the large values at e.g. $k = 301, 401, 501$ mean that the values of $\alpha(r_{i,\ ini}^{dis,\ dir})$ differ strongly between points i of the point distribution. So one can reasonably expect that the sets of scaling indices $\{\alpha_k\}_i,\ k \in [301,\ 600]$ obtained with the discus-like shape are not comparable to template sets.

Low angular resolution of the algorithm

So finally, I made a test concerning the stability of allocation to a certain crystal type depending on the spatial orientation of local structures. I used a point distribution of fcc type with a noise level of only 1% average point distance and rotated it along θ- and

ϕ- directions, each with steps of 10° in the range $[0°, \ldots 170°]$. It turned out that the allocation to a certain crystal type often switches: when rotating a local structure only by 10°, it may switch the allocation type from fcc to hcp or vice versa. Nevertheless, there are some areas of more stability (figure 5.13).

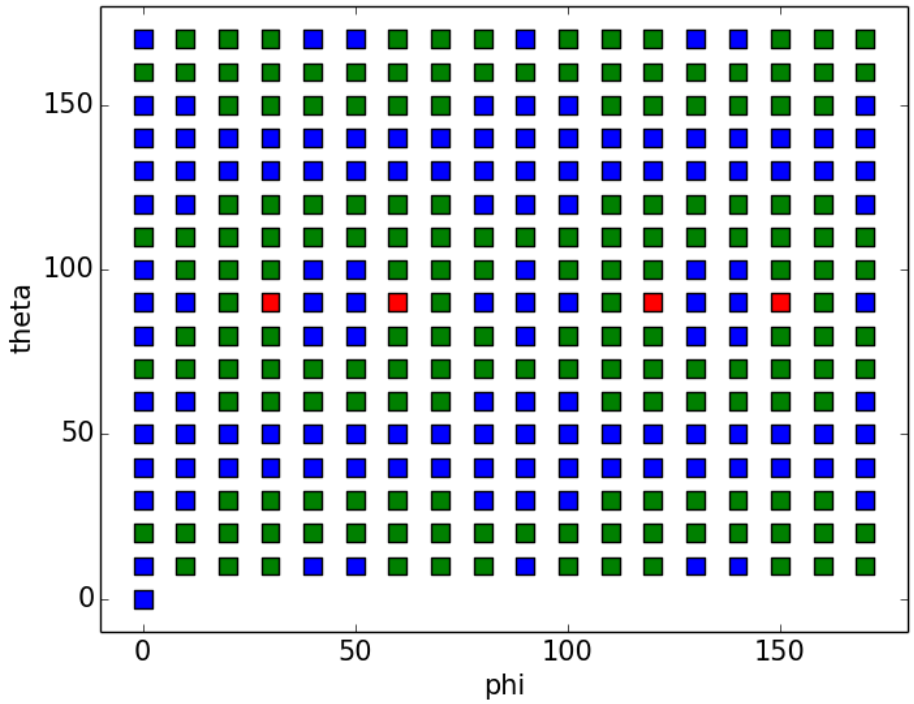

Figure 5.13: The allocation to a certain crystal type bcc, fcc or hcp (especially fcc and hcp) is very sensitive on the angular distribution of points in a local structure. Here rotations of only 10° can alter the allocation type. A blue square means most points of the distribution are allocated to fcc type; a green square means to hcp type; and red means to bcc type. This result is obtained with the same algorithm which has been used in this chapter. The algorithm obviously needs to be improved, since the most information is embedded in the angular distribution of next neighbors in a local structure and less in the scaling.

So finally, it turned out that the usage of the discus-like shape was a disappointing choice. In the following chapter about analyses of 2D point distributions, the scaling indices are calculated along many more different spatial directions using cigar-like (two-dimensional "cigars") shape.

Chapter 6

Scaling indices algorithm and results for 2D point distributions

6.1 Description of algorithm, datasets and results

In the two dimensional datasets only for hexagonal local structures was searched for, since the hexagonal way of stacking has the highest packing density, why two-dimensional distributions usually crystallize in hexagonal shape. The algorithm which was used is basically the same. First local crystalline structures were searched for. These again were defined as clusters of similar structures; then these crystalline structures were allocated to the most similar template. The templates cover different spatial orientations of a hexagonal structure. I used following datasets to test the algorithm:

- Artificially created 2D point distribution of hexagonal type with certain noise levels
- Experimentally obtained dataset showing recrystallization [49]
- Simulated dataset showing the melting of a crystal [50]

With the aid of the artificially created point distributions of hexagons with increasing Gaussian noise level, again the parameters $threshold$, $N_{neighbors}$ for the algorithm were fixed, as described in (4.2). I chose $threshold = 13.6$ and $N_{neighbors} = 18$.

A change in the algorithm for 2D point distributions is the calculation of the scaling indices along many different spatial directions, as described in section 4.3. The aspect ratio again equals $\lambda = 9$. Besides the reference direction 0°, also the directions 10°, 20°, 30°, 40° and 50° were handled. At 60° a hexagonal point distribution is again identical with the reference direction; because the algorithm concentrates on hexagonal structures, no more directions than these six were processed. So in total, I created six templates of six hexagonal point distributions of different orientations, where each template contains the scaling information about six different spatial directions.

Recrystallization dataset

The dataset "recrystallization" is experimentally gained ([49], experiment rI). It contains 2046 2D point distributions / time frames with each almost 1900 particle positions, having a reduced frame rate of 500/3 fps. The experiment was performed in a vacuum chamber, where argon plasma was produced with an asymmetric rf discharge setup. The plasma was kept at room temperature, the peak-to-peak voltage was $-172\ V$ between driven electrode and ground, the lower electrode had a self-bias of $-70\ V$, the neutral gas pressure was 1.938 Pa. Then melamine-formaldehyde particles (diameter $9.19 \pm 0.09\ \mu m$) were interspersed into the chamber, which form a horizontal layer above the electrode. The measurement of the positions was performed by a laser-camera system with top-view. The laser had a power of 132 mW, the camera had a frame rate of 500 fps and a spatial resolution of 0.034 $mmperpx$.

At a certain point of time an electric pulse is emitted to the system "blowing away" the particles. This pulse is created with two parallel wires installed above the electrode, which are connected to the power supply having a voltage of $-253\ V$. The duration of the pulse is controlled by a function generator and equals 0.2 s. After the pulse is switched off the particles begin to crystallize again.

In some figures, I want to illustrate main states of the recrystallization. The plots show the detected crystals, a different color refers to a different orientation from 0° (reference orientation) up to 50°. The non-crystalline structures are marked as grey crosses. Note: only the middle points of the local structures are marked, not the whole structure existing of middle point plus next neighbors. Figure (6.1) shows the results for the first scan of the dataset. The particle layer mainly is homogeneously structured with hexagons with an orientation of 50°. There are areas of different orientations in the corners of the distribution, the transitions can be recognized easily. Points positioned between two areas with different orientations of more than 20° are allocated to non-crystalline by the algorithm, due to the spatial order criterion. Further plots are shown in the appendix (7.27, 7.28, 7.29, 7.30, 7.31). When the pulse is switched on, the particles move too fast to be recognized (7.27). After the pulse is switched off again, they slowly start to arrange, first very disordered (7.28); gradually new crystalline nuclei begin to form and grow (7.29, 7.30). Finally, when the system has cooled down again, the distribution has reached a quite stable shape (last scan: 7.31); the number of clusters with different orientations, which are separated by linear defects, has increased and their positions have changed compared to the first Scan.

Melting dataset

The dataset "melting" was obtained by a simulation performed by Ingo Laut [50]. The dataset consists of 6000 two-dimensional point distributions / time frames with each 378 points. They are a rectangular cut-out from of a circular monolayer point distribution of 16384 points. The simulation was performed similar to the second simulation in [50, p. 5], and figure 4 therein. Their behavior is simulated on the basis of certain equations of

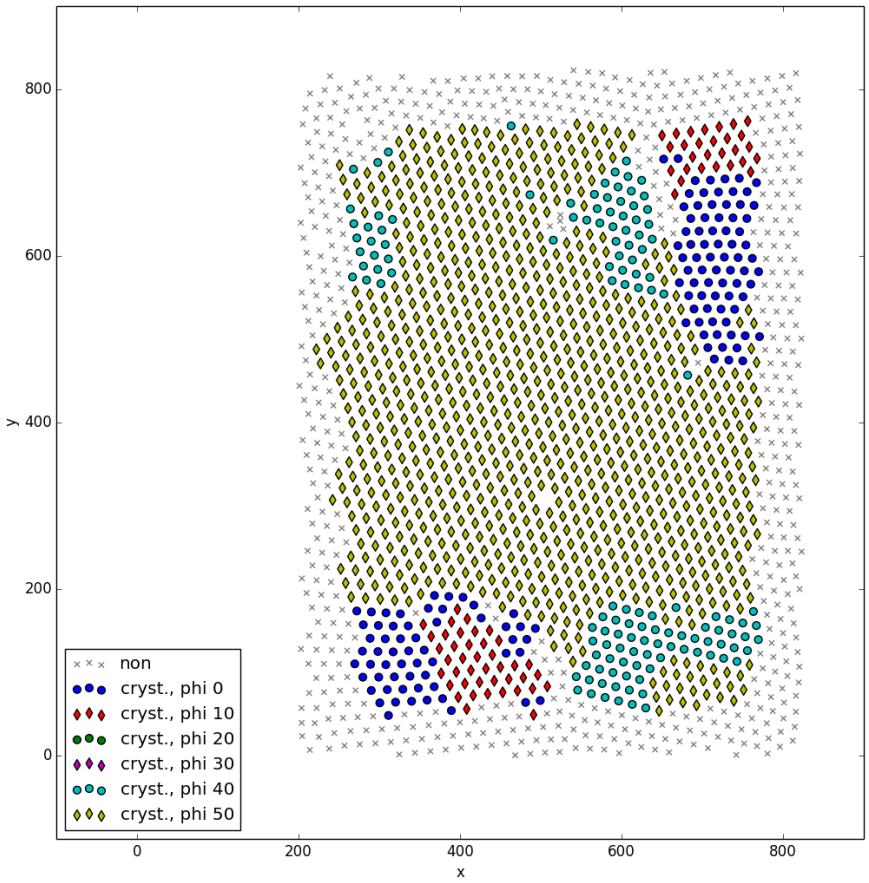

Figure 6.1: This plot shows the visualization of the results of Scan 1 of the recrystallization dataset. Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. Scan 1 is mainly homogeneously oriented, there are some defects in all corners. The defects are usually of linear shape and seperate areas of different orientations.

motion containing forces between the particles (particle charge $Q = 19100\ e$, mass $6.10 \cdot 10^{-13}\ kg$), due to a vertical and an anisotropic horizontal confinement (frequencies are $f_x = 0.180\ Hz,\ f_y = 0.190\ Hz, f_z = 23.00\ Hz$), and due to a Langevin heat bath ($T = 300\ K$). The time step between the frames is $dt = 0.001383956027\ s$. A list of the parameters is given in table (7.1) in the appendix. The anisotropy in the horizontal confinement leads to synchronization effects of the particles and an onset of motion intensifying over time. The result is the melting of the crystals, which is visualized in following plots using the scaling indices algorithm.

Figure (6.2) shows scan 3000, so the middle of the run. The motion of the particles still is very tiny, so the distribution is very homogeneously shaped and oriented. Till about scan 4700 (figure 7.32 appendix), the motion of the particles intensifies, which behave like several strings oscillating against each other, still without the crystal loosing its overall hexagonal homogeneity. The crystal at this point of time only is melting from the border introversively. Figure (7.33 appendix) shows scan 5000. Between scans 4700 and 5000 the motion has strongly intensified, which has leaded to a partial melting of the crystal. The remaining crystalline clusters are mostly of the original orientation (0°) and of 10° clockwise and counter-clockwise rotations (orientations 10° and 50°). Till the end the fluctuations are quite strong (scan 5400 7.34 appendix), but at the very end the size of the crystalline cluster even has increased a bit again (scan 5999 7.35 appendix). The recrystallization at the end of the run would not have happened if the dataset was experimentally gained; in experiments the synchronization would get lost and the crystal would melt completely. The simulation, however, does not model this; here the synchronization leads to a oscillation between melting and recrystallization (see [50]). The graph in figure (6.3) exhibits the time evolution of the absolute number of detected crystalline structures itemized by their orientation, starting at scan 3000.

6.2 Detection of defect lines

The criterion of stability which I chose, namely the requirement of a similar neighborhood for defining a local structure as crystalline, has the convenient effect that transitions between different crystalline clusters - as long as they are rough enough - are detected. “Detected” means the local structures at those positions are allocated to “non allocated”. This effect is controlled by two parameters: the similarity *threshold* and $N_{neighbors}$, see section 4.2. Because *threshold* is fixed by the criterion described in (4.2) and because $N_{neighbors}$ has very low effect on keeping this criterion, $N_{neighbors}$ can be varied in some range (the parameter *threshold* nevertheless is adapted slightly). Furthermore it is possible just to forget about the criterion and to decrease *threshold* in order to detect edges. With smaller *threshold*, local structures which were used to be detected as crystalline transitions between larger crystalline clusters of different orientations, eventually are detected as edges. With the normal, non-lowered value of *threshold*, transitions between crystalline clusters with 10° difference in their orientation are detected as smooth transitions; transitions of

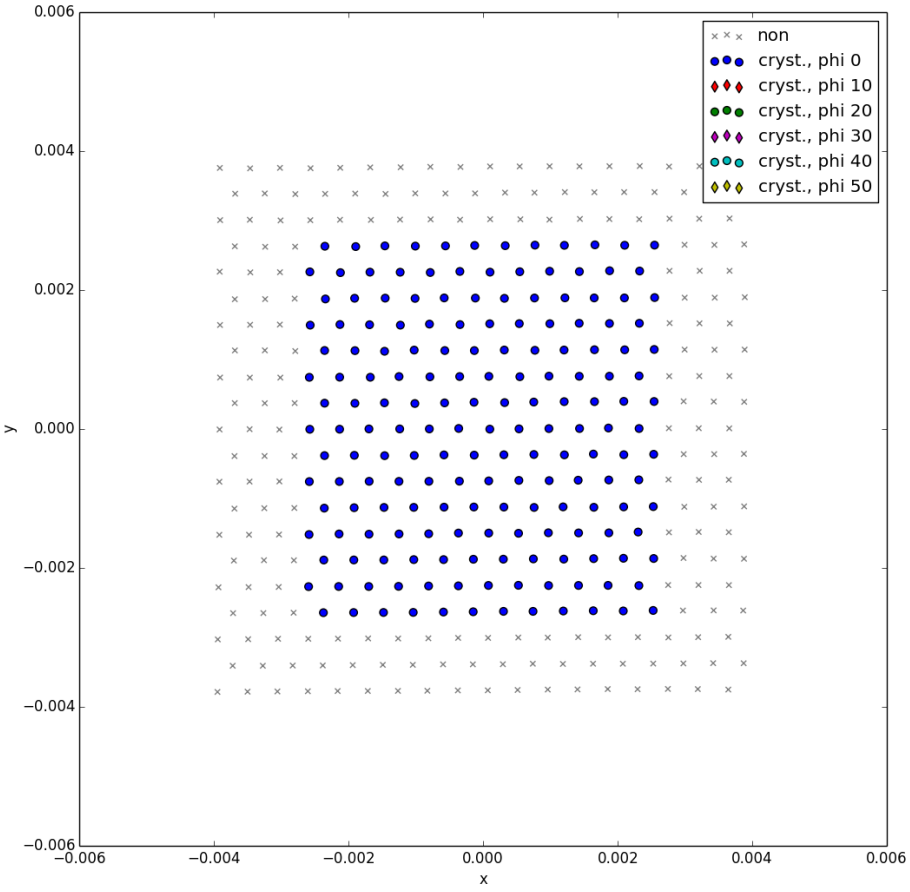

Figure 6.2: This plot shows the visualization of the results of Scan 3000 of the melt dataset. Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. Scan 3000 is nicely homogeneously and hexagonally structured and oriented. The movement of the points is very tiny at this point of time.

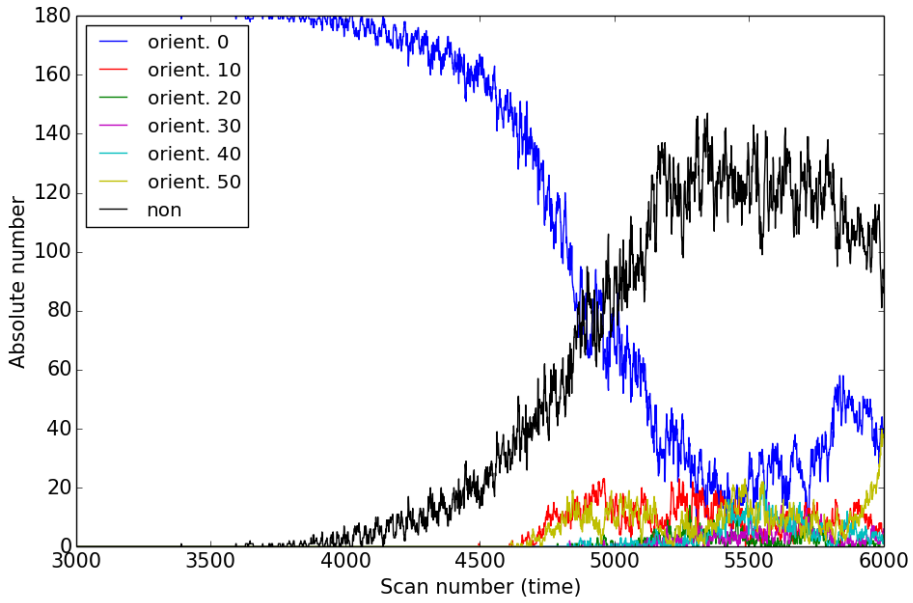

Figure 6.3: The number of crystalline structures stays constant during the first half of the run (till scan 3000). Then, in a oscillating way, the number decreases. Till about scan 4700, the crystal melts only in the border region, without taking another orientation. Then, also the inner regions of the crystal start to change, melting instantly or by taking another orientation (10° and 50°). The oscillating way of the decreasing number of crystals comes from the kind of motion of the particles. Due to the synchronization effects they oscillate as strings against each other, sometimes having a more orderly shape, sometimes a less orderly one. The frequency of this oscillation is about 1 per 20 time frames, which was estimated from the graph and checked with a power spectrum (figure 7.36 appendix). Around scan 5500, the fluctuations are very strong, but at their maximum still about 20% of the points are allocated to a crystalline category. At the very end, this number even increases again to more than 40%. So the crystal has not melted completely, but the synchronization is still present and leads to recrystallization. Indeed, as described in [50], the simulation creates oscillating alternations between partial melting and recrystallization.

20° and 30° are detected as edges.

In the appendix, you can find some figures which demonstrate the effect of choosing different values of *threshold* and $N_{neighbors}$ on the detection of defect lines. In figure (7.37) the analysis was done with $N_{neighbors} = 6$, in figure (7.38) with $N_{neighbors} = 18$. The increasing of this parameter strengthens the upper and middle defect line, but still transitions between areas with 10° difference in their orientation are not taken as edge. In the figures (7.39) and (7.40) $N_{neighbors} = 18$ is kept, but *threshold* is lowered each time. Now structures, which used to be transitions, e.g. the transition in the above quarter between violet and yellow diamonds (orientations 30° and 50°, respectively) are allocated to non-crystalline.

Chapter 7

Summary

Scaling indices, whose usage originates from the time series analysis of chaotic systems, are a local measure for the structure of a point distribution (see section 1.2 and [20]). Because the phase space trajectories of chaotic systems settle down on strange attractors with fractal geometries, new methods were needed for estimating the fractal dimensionality of these attractors (for an introduction to chaos and fractals read e.g. Strogatz [16]). Scaling indices measure the local property of a point in a distribution how its number of next neighbors scales with a radius-like parameter. This basic algorithm is extended by including all points in the distribution and weighting their impact by their distance ("weighted scaling indices"). This offers a nice advantage compared with the bond order parameters method [11]: a definition of "next neighbors" is not necessary, making the scaling indices algorithm regarding this more stable against noisy fluctuations of the particle positions. The angular distribution of the neighbors, which often is anisotropic, then is taken into account by introducing an anisotropic distance measure. Finally, the numerical algorithm calculates a set of scaling indices for each data point; these sets are described as points in the parameter space, called α-space (section 1.5). Single crystalline structures then are classified by comparing their sets of scaling indices to sets of templates of crystalline structures (section 4.3).

The method of scaling indices has proven to be useful for the discrimination of crystal types, checked by means of artificial crystalline structures. It even has shown a better quality of discrimination than the method of bond order parameters, in terms of a higher robustness against the noise level of the point distributions (chapter 3). This result was encouraging to follow up on the improvement of the algorithm. For analyzing unknown data, the algorithm was made scaling-free and a criterion of spatial order was introduced (chapter 4). This criterion ensures that high-noise structures are rejected in the crystal-detection process and only larger clusters of similar structures are classified as crystalline. It automatically brings along the detection of defect lines in point distributions (section 6.2), as well as the detection of the border region. Furthermore, two different kinds of shapes of the cumulative distribution function were tried out. The decision to use the discus-like shape, however, in the end was disappointing (section 5.5). Although the results for the PK-3-Plus dataset were reliable (section 5.3), the result for the data from PKE-

Nefedov strongly differed compared with the modified bond order method (section 5.4). A principal component analysis of the scaling indices results and a further test of the rotational stability of the allocation process revealed the non-suitability of the discus-like shape. So the increase of the angular resolution of the algorithm was followed up on. This was tested by two-dimensional point distributions (chapter 6), having led to very nice results differing hexagonal structures of several orientations. The feature of the scaling indices algorithm of being able to detect different orientations of crystalline structures is a direct consequence of not being rotationally invariant. Although this might be seen as disadvantage due to a much larger number of templates and longer computing times, the advantage of having gained more information about the structure of the data predominates.

By increasing the angular resolution of the algorithm for 3D distributions, as good results can be suggested. After having achieved this, a combination of the allocation algorithm with tracking of the particles is the next logical step. So a time series of the allocations can be created; furthermore, if the time resolution is high enough, a time stability criterion for defining crystals can be introduced. The time series then enables the analysis of transitions between crystal types, especially during the nucleation and creation process of a crystalline cluster.

Bibliography

[1] C. K. Goertz, "Dusty plasmas in the solar system," *Rev. Geophys.*, vol. 27, no. 2, pp. 271–292, 1989.

[2] I. Mann, N. Meyer-Vernet, and A. Czechowski, "Dust in the planetary system: Dust interactions in space plasmas of the solar system," *Physics Reports*, vol. 536, no. 1, pp. 1–39, 2014.

[3] H. Izeki, "Coulomb solid of small particles in plasmas," *Phys. Fluids*, vol. 29, p. 1764, 1986.

[4] H. Thomas, G. E. Morfill, V. Demmel, J. Goree, B. Feuerbacher, and D. Möhlmann, "Plasma crystal: Coulomb Crystallization in a Dusty Plasma," *Phys. Rev. Lett.*, vol. 73, pp. 652–655, 1994.

[5] H. M. Thomas and G. E. Morfill, "Melting dynamics of a plasma crystal," *Nature*, vol. 379, no. 6568, pp. 806–809, 1996.

[6] S. A. Khrapak, B. A. Klumov, P. Huber, V. I. Molotkov, A. M. Lipaev, V. N. Naumkin, A. V. Ivlev, H. M. Thomas, M. Schwabe, G. E. Morfill, O. F. Petrov, V. E. Fortov, Y. Malentschenko, and S. Volkov, "Fluid-solid phase transitions in three-dimensional complex plasmas under microgravity conditions," *Phys. Rev. E*, vol. 85, no. 066407, 2012.

[7] M. Rubin-Zuzic, G. E. Morfill, A. V. Ivlev, R. Pompl, B. A. Klumov, W. Bunk, H. M. Thomas, H. Rothermel, O. Havnes, and A. Fouquet, "Kinetic development of crystallization fronts in complex plasmas," *Nat. Phys.*, vol. 2, pp. 181–185, 2006.

[8] S. A. Khrapak, B. A. Klumov, P. Huber, V. I. Molotkov, A. M. Lipaev, V. N. Naumkin, H. M. Thomas, A. V. Ivlev, G. E. Morfill, O. F. Petrov, V. E. Fortov, Y. Malentschenko, and S. Volkov, "Freezing and Melting of 3D Complex Plasma Structures under Microgravity Conditions Driven by Neutral Gas Pressure Manipulation," *Phys. Rev. Lett.*, vol. 106, no. 20, p. 205001, 2011.

[9] K. Michielsen and H. D. Raedt, "Integral-geometry morphological image analysis," *Physics Reports*, vol. 347, no. 6, pp. 461 – 538, 2001.

[10] G. E. Schröder-Turk, W. Mickel, S. C. Kapfer, F. M. Schaller, B. Breidenbach, D. Hug, and K. Mecke, "Minkowski tensors of anisotropic spatial structure," *New J. Phys.*, vol. 15, no. 8, p. 083028, 2013.

[11] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Phys. Rev. B*, vol. 28, no. 784, 1983.

[12] H. Whitney, "Differentiable manifolds," *Ann. Math.*, vol. 37, pp. 645–680, 1936.

[13] F. Takens, *Lecture notes in Mathematics vol. 898: Dynamical Systems and Turbulence, Warwick 1980*, ch. Detecting strange attractors in turbulence. Rand, D. A. and Young, L.-S., Springer Verlag, 1981.

[14] A. Fraser and H. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, p. 1134, 1986.

[15] M. Kennel, R. Brown, and H. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, no. 6, pp. 3403–3411, 1992.

[16] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Addison-Wesley, 1994.

[17] E. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, pp. 130–141, 1963.

[18] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications, Second Edition*. John Wiley & Sons, 2003.

[19] P. Grassberger and I. Procaccia, "Characterization of Strange Attractors," *Phys. Rev. Lett.*, vol. 50, no. 5, pp. 346–349, 1983.

[20] C. Räth, R. Monetti, J. Bauer, I. Sidorenko, D. Müller, M. Matsuura, E.-M. Lochmüller, P. Zysset, and F. Eckstein, "Strength through structure: visualization and local assessment of the trabecular bone structure," *New J. Phys.*, vol. 10, no. 125010, 2008.

[21] K. Pearson, "On lines and planes of closest fit to a system of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 6, no. 2, pp. 559–572, 1901.

[22] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Icosahedral Bond Orientational Order in Supercooled Liquids," *Phys. Rev. Lett.*, vol. 47, pp. 1297–1300, 1981.

[23] P. Rein ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, "Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling," *The Journal of Chemical Physics*, vol. 104, no. 24, pp. 9932–9947, 1996.

[24] J. R. Errington, P. G. Debenedetti, and S. Torquato, "Quantification of order in the Lennard-Jones system," *The Journal of Chemical Physics*, vol. 118, no. 5, pp. 2256–2263, 2003.

[25] M. D. Rintoul and S. Torquato, "Computer simulations of dense hardsphere systems," *The Journal of Chemical Physics*, vol. 105, no. 20, pp. 9258–9265, 1996.

[26] S. Torquato, T. M. Truskett, and P. G. Debenedetti, "Is Random Close Packing of Spheres Well Defined?," *Phys. Rev. Lett.*, vol. 84, pp. 2064–2067, 2000.

[27] P. Richard, A. Gervois, L. Oger, and J.-P. Troadec, "Order and disorder in hard-sphere packings," *EPL*, vol. 48, no. 4, p. 415, 1999.

[28] Y. Jin and H. A. Makse, "A first-order phase transition defines the random close packing of hard spheres," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 23, pp. 5362 – 5379, 2010.

[29] B. A. Klumov, S. A. Khrapak, and G. E. Morfill, "Structural properties of dense hard sphere systems near random close packing," *AIP Conference Proceedings*, vol. 1397, no. 1, pp. 441–442, 2011.

[30] B. A. Klumov, S. A. Khrapak, and G. E. Morfill, "Structural properties of dense hard sphere packings," *Phys. Rev. B*, vol. 83, p. 184105, 2011.

[31] U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, and D. A. Weitz, "Real-Space Imaging of Nucleation and Growth in Colloidal Crystallization," *Science*, vol. 292, no. 5515, pp. 258–262, 2001.

[32] S. Auer and D. Frenkel, "Numerical prediction of absolute crystallization rates in hard-sphere colloids," *The Journal of Chemical Physics*, vol. 120, no. 6, pp. 3015–3029, 2004.

[33] S. Mitic, B. A. Klumov, U. Konopka, M. H. Thoma, and G. E. Morfill, "Structural Properties of Complex Plasmas in a Homogeneous dc Discharge," *Phys. Rev. Lett.*, vol. 101, p. 125002, 2008.

[34] B. Klumov, M. Rubin-Zuzic, and G. Morfill, "Crystallization waves in a dusty plasma," *JETP Letters*, vol. 84, no. 10, pp. 542–546, 2007.

[35] B. Klumov and G. Morfill, "Characteristics of crystallization of complex plasmas in narrow channels," *Journal of Experimental and Theoretical Physics*, vol. 107, no. 5, pp. 908–915, 2008.

[36] B. Klumov, P. Huber, S. Vladimirov, H. Thomas, A. Ivlev, G. Morfill, V. Fortov, A. Lipaev, and V. Molotkov, "Structural properties of 3D complex plasmas: experiments versus simulations," *Plasma Physics and Controlled Fusion*, vol. 51, no. 12, p. 124028, 2009.

[37] B. Klumov and G. Morfill, "Complex plasma in narrow channels: Impact of confinement on the local order," *JETP Letters*, vol. 85, no. 10, pp. 498–502, 2007.

[38] B. Klumov and G. Morfill, "Effect of confinement on the crystallization of a dusty plasma in narrow channels," *JETP Letters*, vol. 87, no. 8, pp. 409–413, 2008.

[39] B. Zoetekouw, "The body centered cubic crystal structure." https://commons.wikimedia.org/wiki/File%3ALattice_body_centered_cubic.svg. Content of Wikipedia.

[40] B. Zoetekouw, "The face centered cubic crystal structure." https://commons.wikimedia.org/wiki/File%3ALattice_face_centered_cubic.svg. Content of Wikipedia.

[41] Dornelf, "Hexagonal close packed crystal structure." https://commons.wikimedia.org/wiki/File%3AHexagonal_close_packed.svg. Content of Wikipedia.

[42] F. A. Lindemann, "Über die Berechnung molekularer Eigenfrequenzen," *Physikalische Zeitschrift*, vol. 11, pp. 609–612, 1910.

[43] H. M. Thomas, G. E. Morfill, V. E. Fortov, A. V. Ivlev, V. I. Molotkov, A. M. Lipaev, T. Hagl, H. Rothermel, S. A. Khrapak, R. K. Suetterlin, M. Rubin-Zuzic, O. F. Petrov, V. I. Tokarev, and S. K. Krikalev, "Complex plasma laboratory PK-3 Plus on the International Space Station," *New J. Phys.*, vol. 10, no. 033036, 2008.

[44] P. Huber, "Untersuchung von 3-D Plasmakristallen," *Ludwig-Maximilians-Universität München*, 2011.

[45] A. P. Nefedov, G. E. Morfill, V. E. Fortov, H. M. Thomas, H. Rothermel, T. Hagl, A. V. Ivlev, M. Zuzic, B. A. Klumov, A. M. Lipaev, V. I. Molotkov, O. F. Petrov, Y. P. Gidzenko, S. K. Krikalev, W. Shepherd, A. I. Ivanov, M. Roth, H. Binnenbruck, J. A. Goree, and Y. P. Semenov, "PKE-Nefedov: plasma crystal experiments on the International Space Station," *New J. Phys.*, vol. 5, no. 1, p. 33, 2003.

[46] C. Dietz, "Experimental Investigations and Imageprocessing of Crystalline Structures in Complex Plasmas," *Justus-Liebig-Universität Giessen*, 2015.

[47] C. Dietz, "Pressure Dependence and Structural Analysis of Plasma Crystals under Laboratory Conditions," *Workshop Diagnostics and Simulations of Dusty Plasmas at Christian-Albrechts-Universität Kiel*, 2015.

[48] J. D. H. Leitte, "Vorlesung Algorithmische Geometrie: 8. Voronoidiagramme und Delaunaytriangulierung," *Computer Graphics and Visualization group, Ruprecht-Karls-Universität Heidelberg, http://www.iwr.uni-heidelberg.de/groups/CoVis/Teaching/AG_SS12/AG_8_Voronoi.pdf*, 2012.

[49] C. Knapek, "Phase Transitions in Two-Dimensional Complex Plasmas," *Ludwig-Maximilians-Universität München*, 2010.

[50] I. Laut, C. Räth, S. Zhdanov, V. Nosenko, L. Couëdel, and H. M. Thomas, "Synchronization of particle motion in compressed two-dimensional plasma crystals," *EPL*, vol. 110, no. 65001, 2015.

Appendix

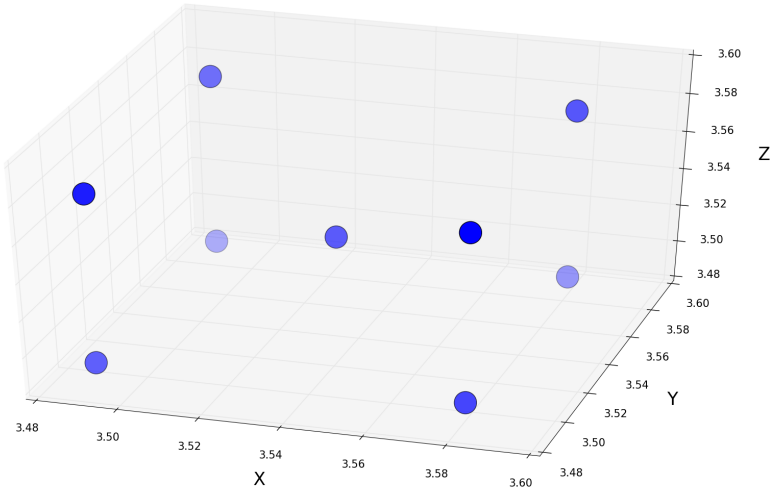

Figure 7.1: Unit cell of the crystal type bcc, taken from artificially created point distributions used in this thesis. The unit cell contains 8 points in the corners and 1 in the middle of its body. It also shows the coordination of the crystal type bcc: the coordination number is 8.

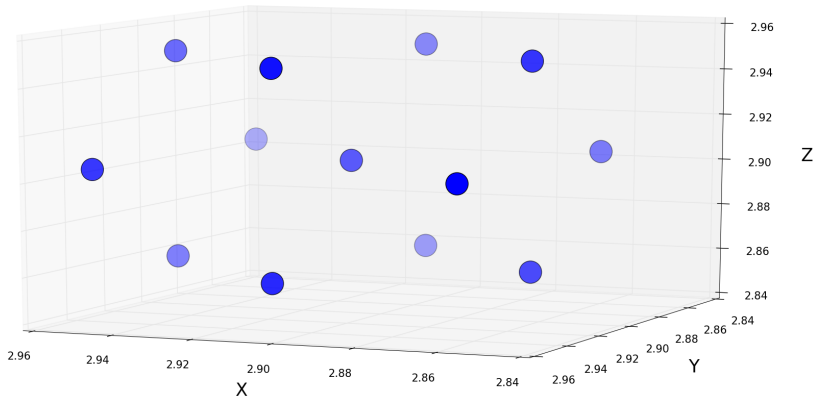

Figure 7.2: Cut-out from the artificially created point distribution of crystal type fcc. Here the coordination of the fcc type is shown: the coordination number is 12. The stacking order $ABC\ldots$ can be recognized (see also figure 7.4): the top right and bottom left layer each containing 3 points have different order.

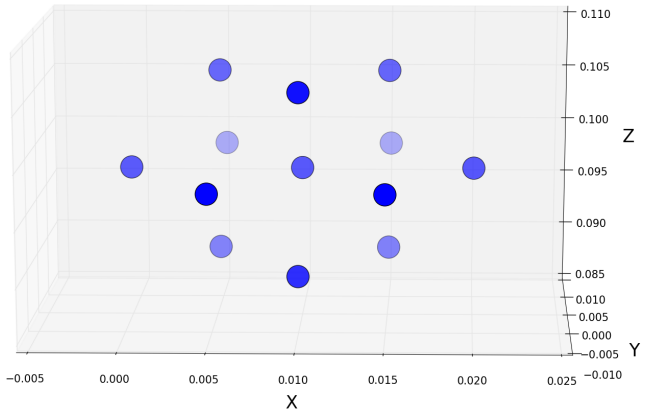

Figure 7.3: Cut-out from the artificially created point distribution of crystal type hcp. The hcp type is characterized by this stacking order $ABAB\ldots$ (see figure 7.4). The coordination number is 12.

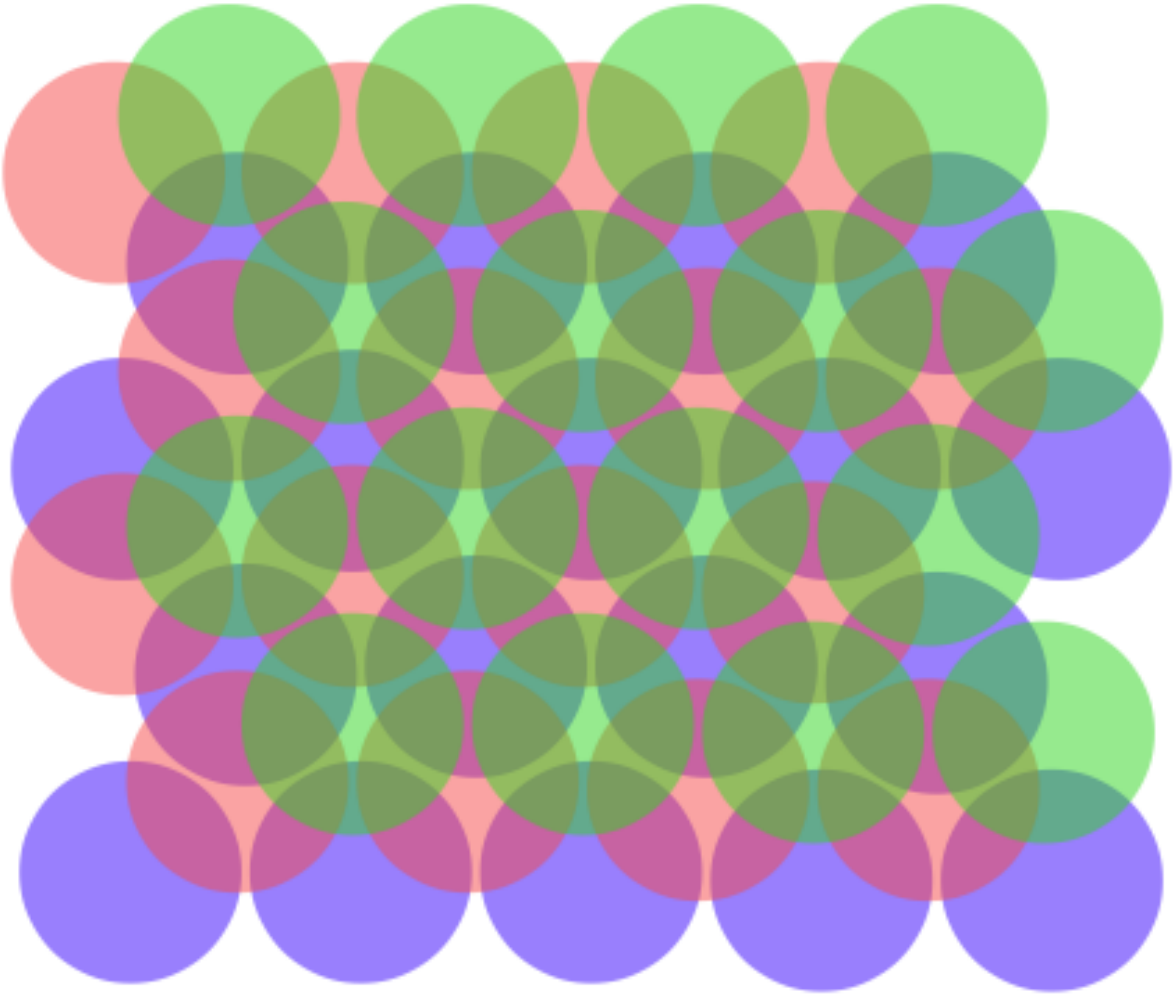

Figure 7.4: For stacking balls, the close-packing is an intuitive and efficient way. The balls of the first layer (blue) are stacked in hexagonal way (stacking order A). The balls of the second layer then are placed inside the hallows of the first layer, accordingly at the red or green positions (order B or C, respectively). For the repetition of the stacking orders $ABCABC\ldots$ the cubic close-packed type (or "face-centered cubic" fcc) is formed, for $ABAB\ldots$ the hexagonal close-packed type (hcp).

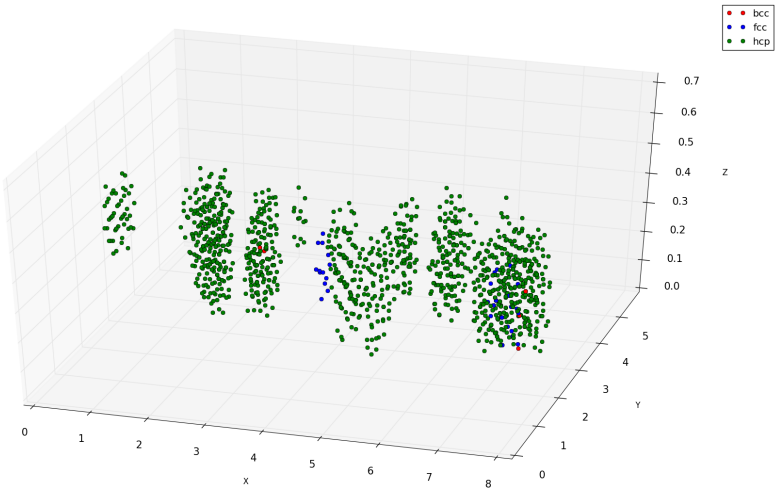

Figure 7.5: Illustration of **Scan 1** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

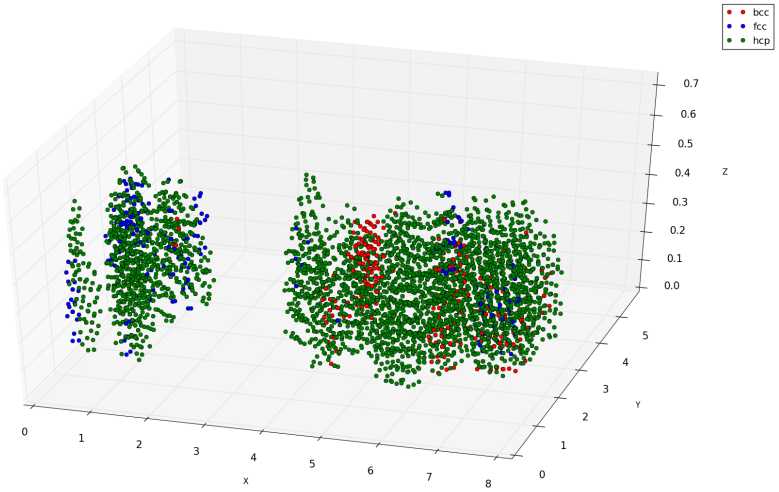

Figure 7.6: Illustration of **Scan 2** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

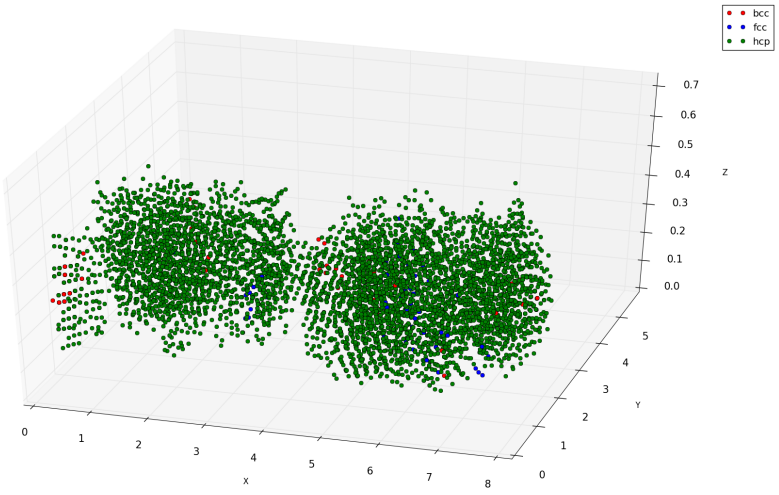

Figure 7.7: Illustration of **Scan 3** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

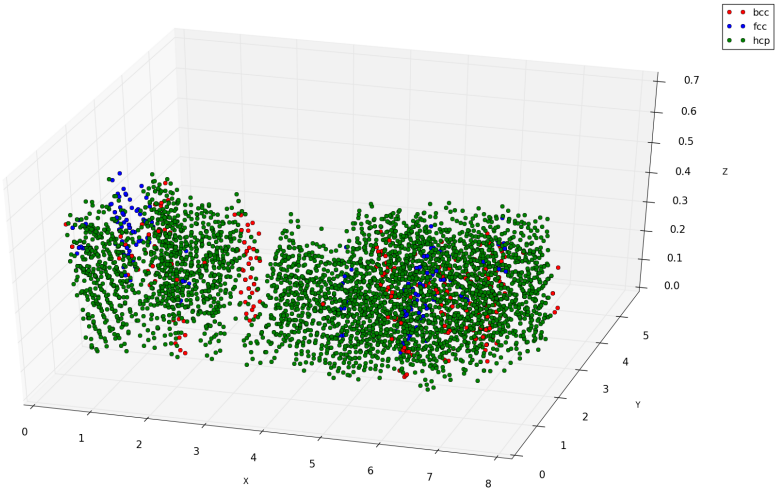

Figure 7.8: Illustration of **Scan 4** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

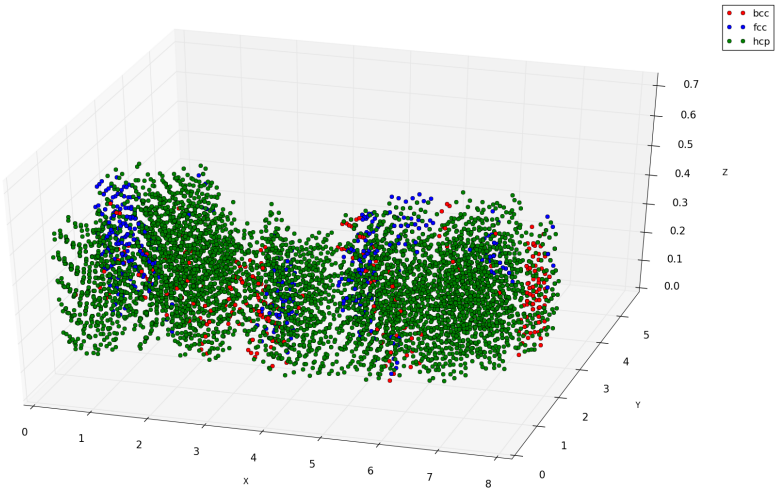

Figure 7.9: Illustration of **Scan 5** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

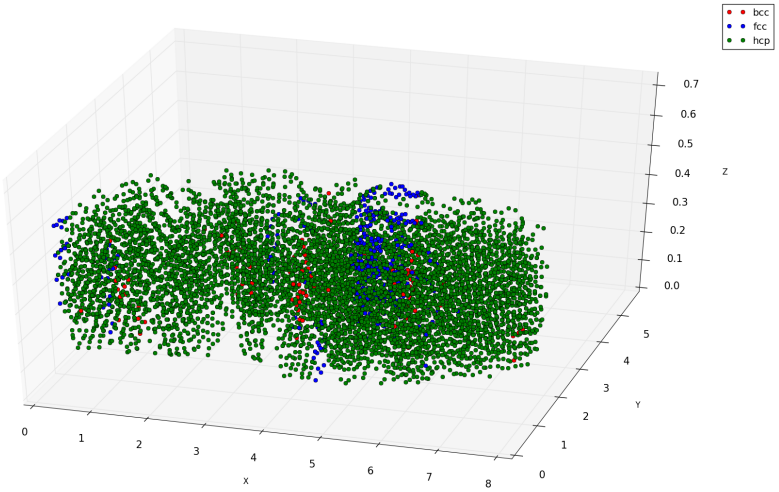

Figure 7.10: Illustration of **Scan 6** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

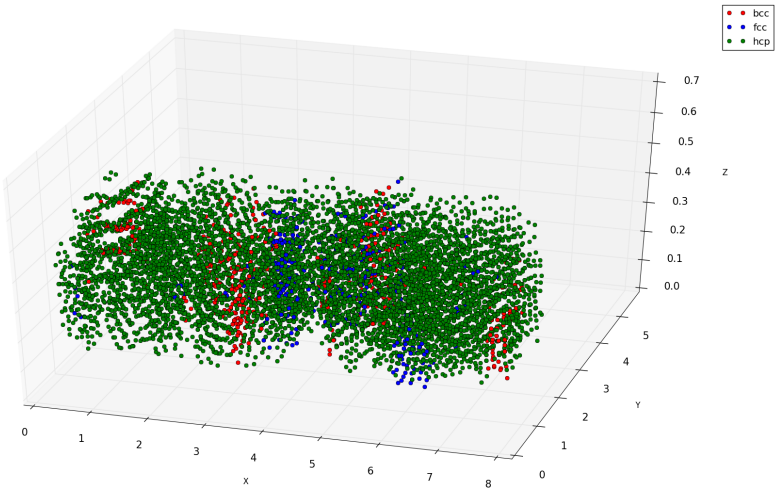

Figure 7.11: Illustration of **Scan 7** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

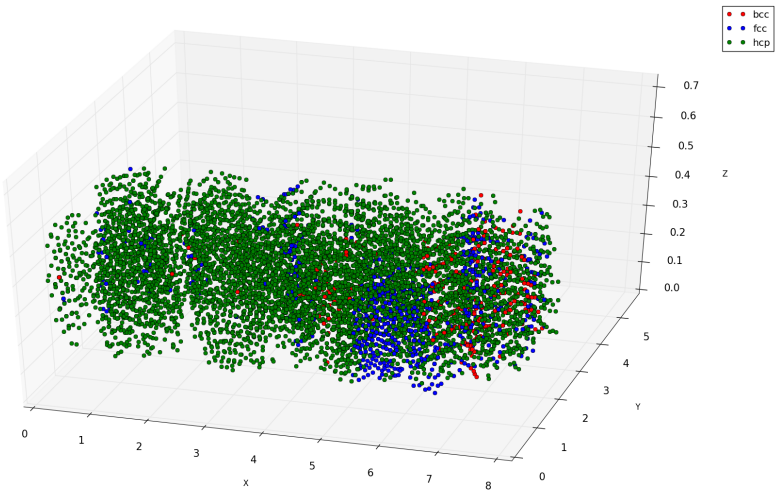

Figure 7.12: Illustration of **Scan 8** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

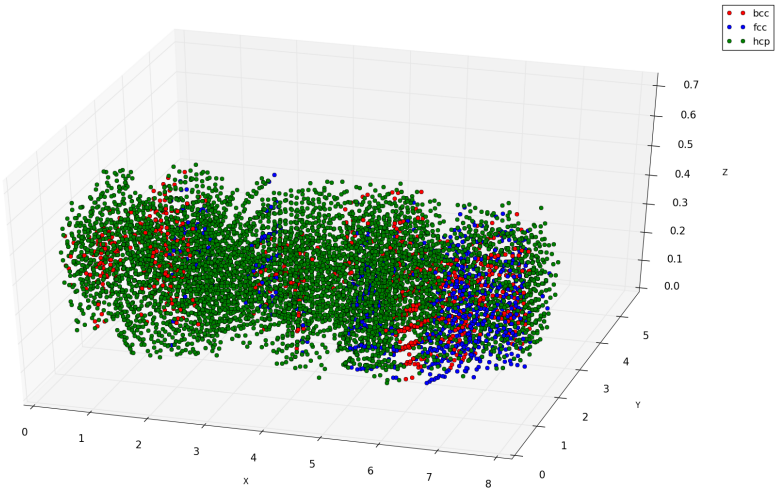

Figure 7.13: Illustration of **Scan 9** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

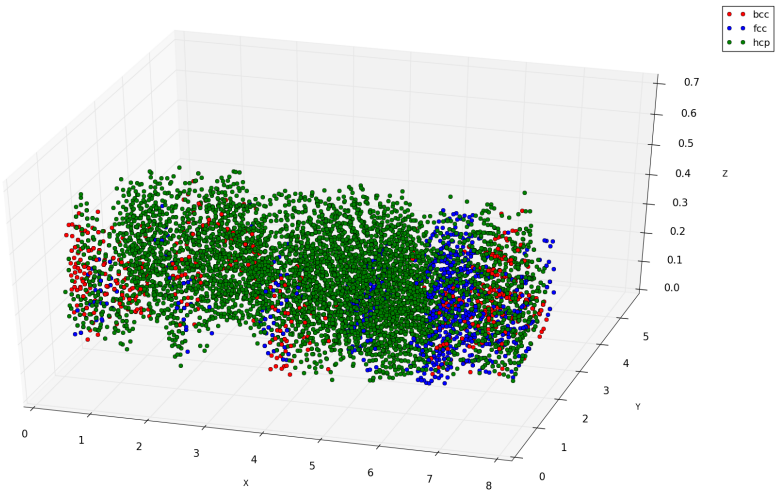

Figure 7.14: Illustration of **Scan 10** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

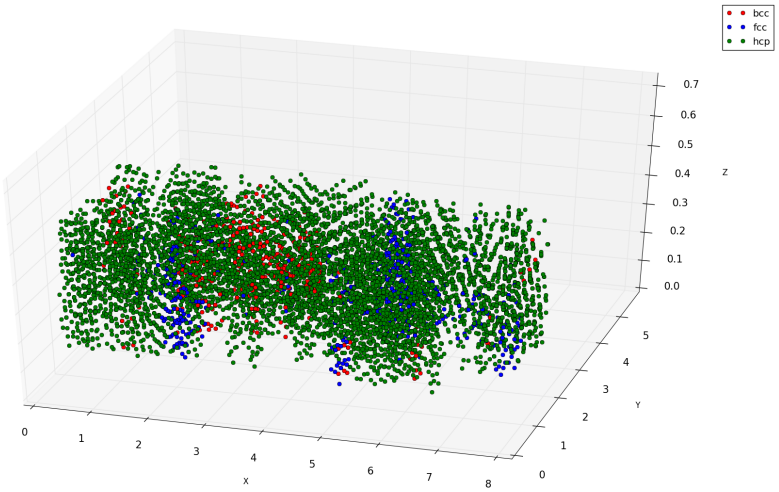

Figure 7.15: Illustration of **Scan 11** of the PK-3-Plus dataset. Only crystalline structures detected using the **scaling indices** method are shown. Red circles refer to bcc type, blue to fcc and green to hcp. Plotted is the middle point of a crystalline structure as well as its 8 next neighbors for bcc, or 12 next neighbors for fcc and hcp.

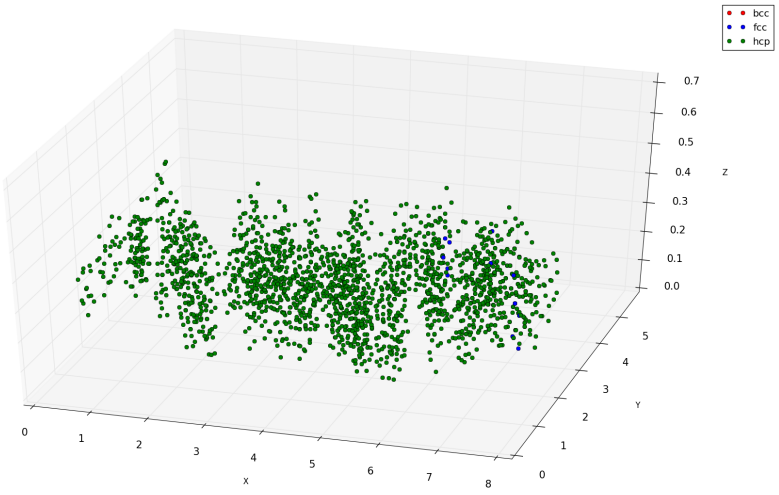

Figure 7.16: Illustration of **Scan 1** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

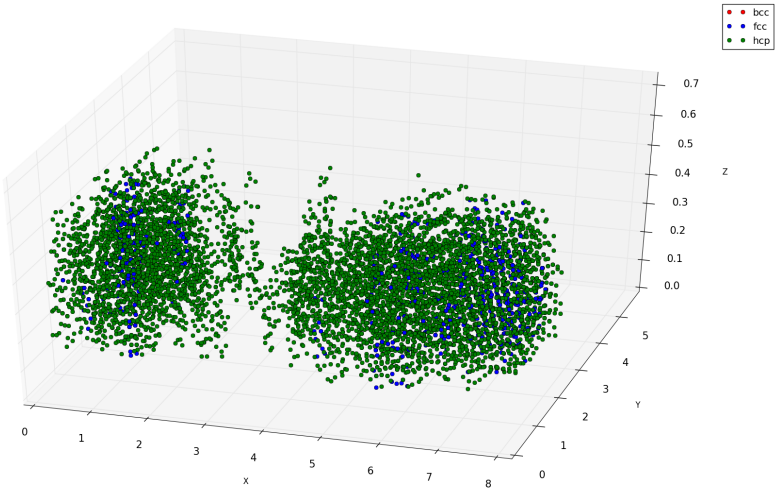

Figure 7.17: Illustration of **Scan 2** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

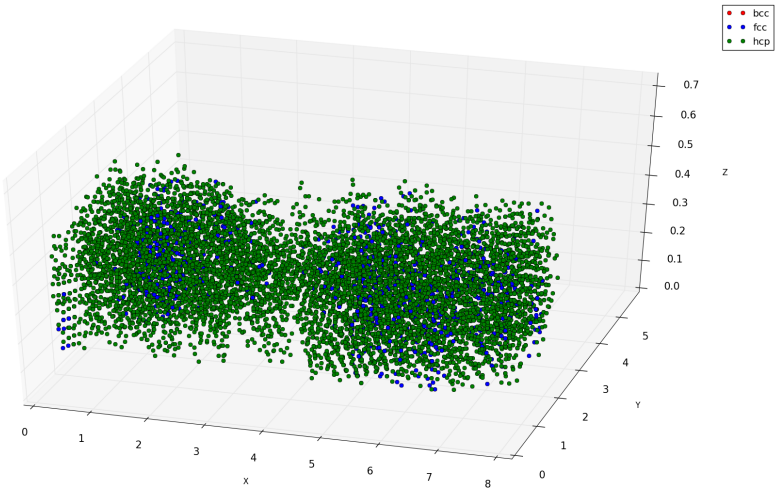

Figure 7.18: Illustration of **Scan 3** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

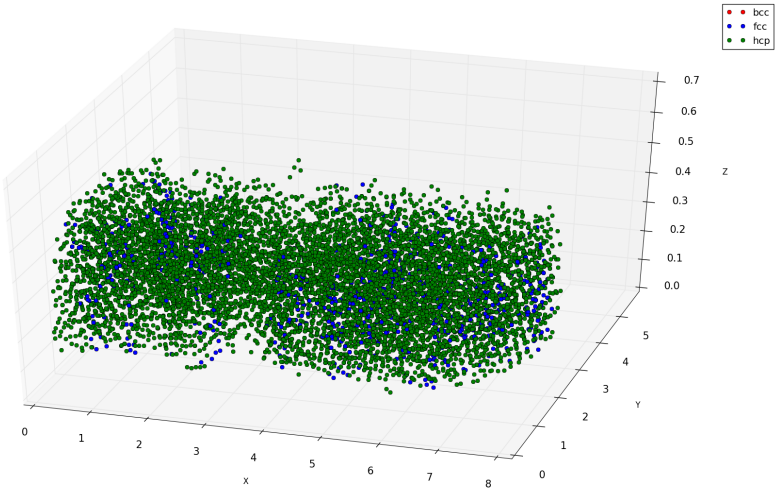

Figure 7.19: Illustration of **Scan 4** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

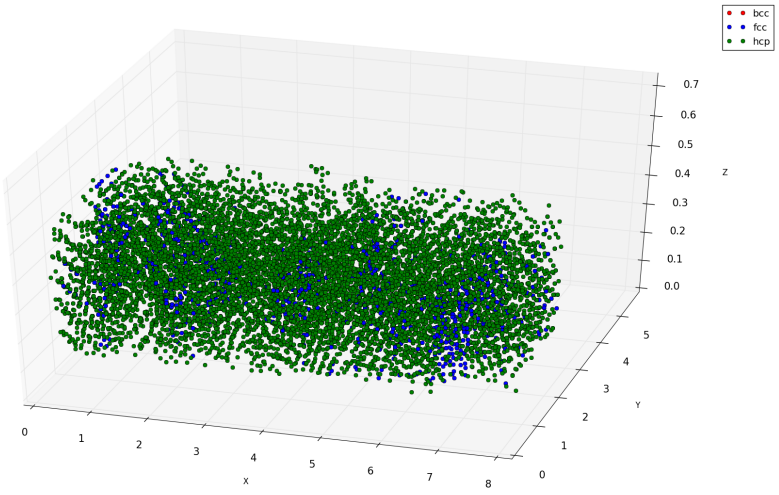

Figure 7.20: Illustration of **Scan 5** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

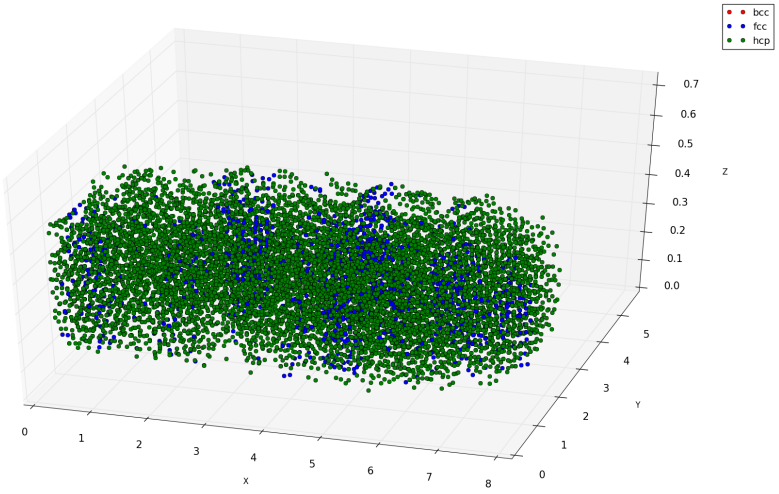

Figure 7.21: Illustration of **Scan 6** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

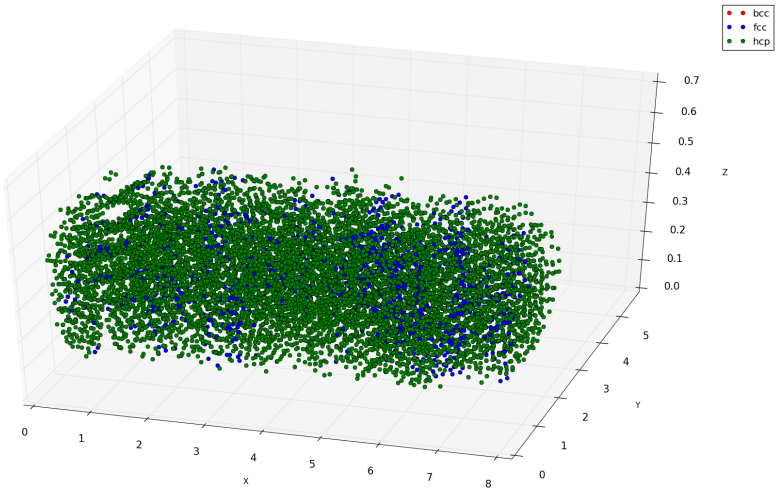

Figure 7.22: Illustration of **Scan 7** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

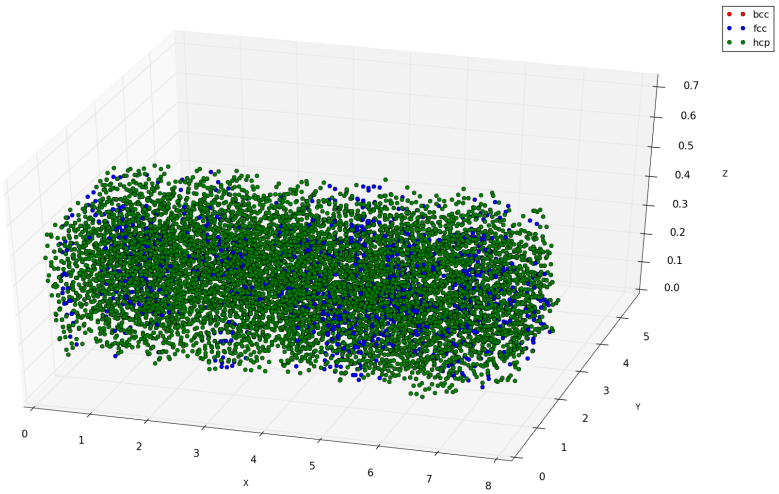

Figure 7.23: Illustration of **Scan 8** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

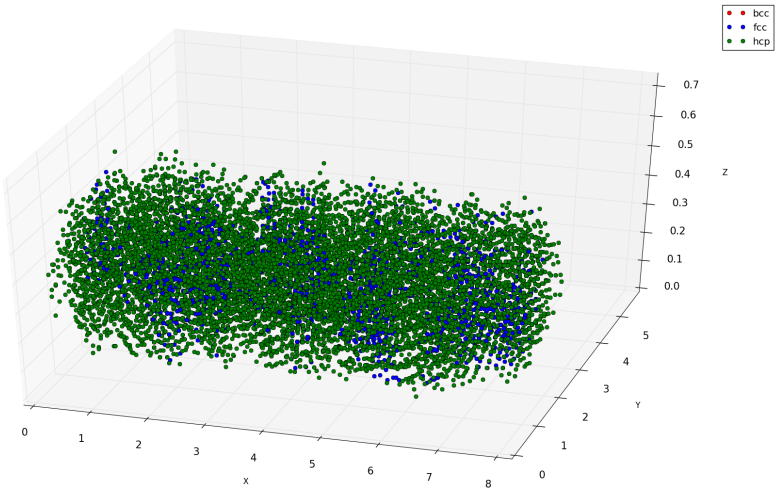

Figure 7.24: Illustration of **Scan 9** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

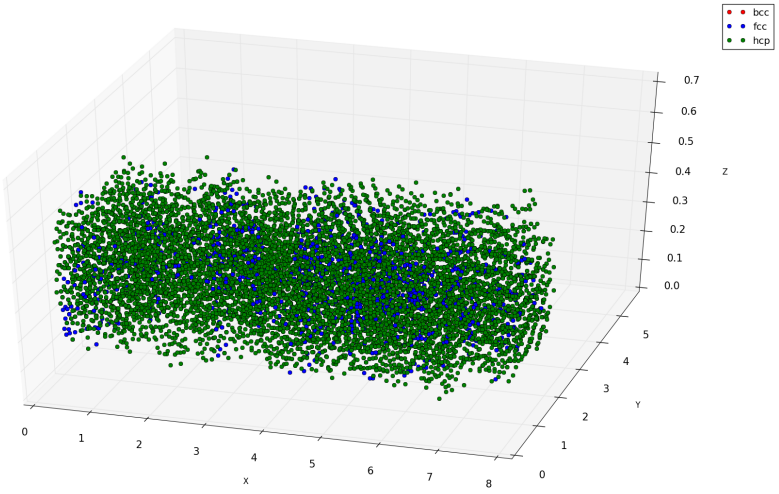

Figure 7.25: Illustration of **Scan 10** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

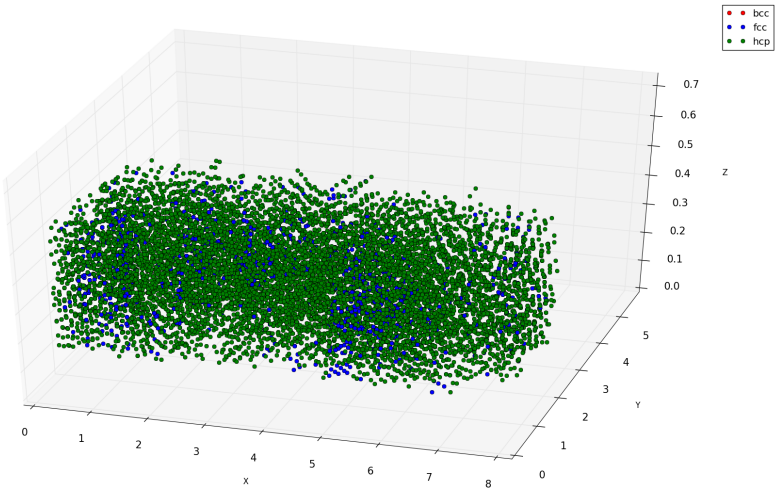

Figure 7.26: Illustration of **Scan 11** of the PK-3-Plus dataset. Only crystalline structures detected using the **bond order parameters** method are shown. Blue circles refer to fcc type, green circles to hcp. Plotted is the middle point of a crystalline structure as well as its 12 next neighbors.

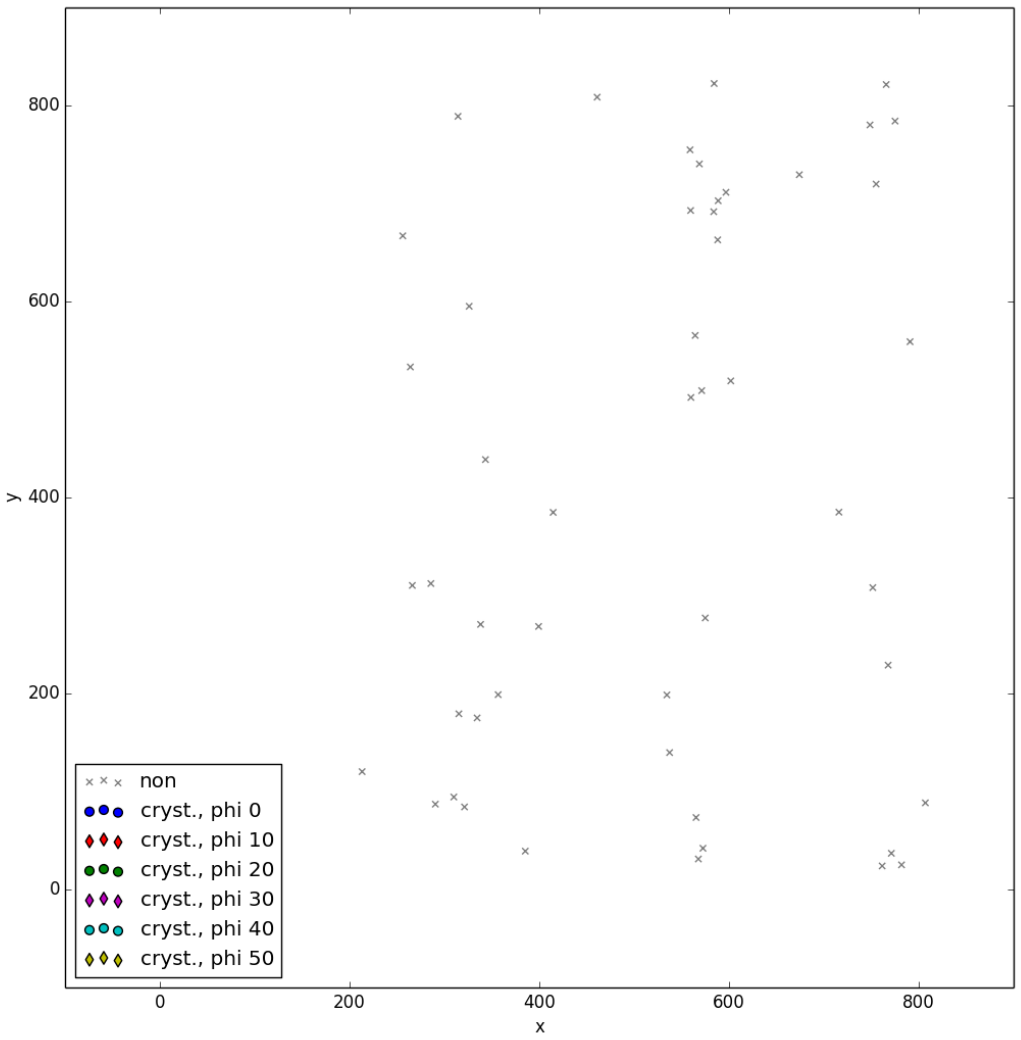

Figure 7.27: Recrystallization dataset, scan 450: when the electric pulse is switched on and blows away the particles, they move too fast for being recognized. The electric pulse has a duration of 0.2 s, this corresponds to about 30 frames.

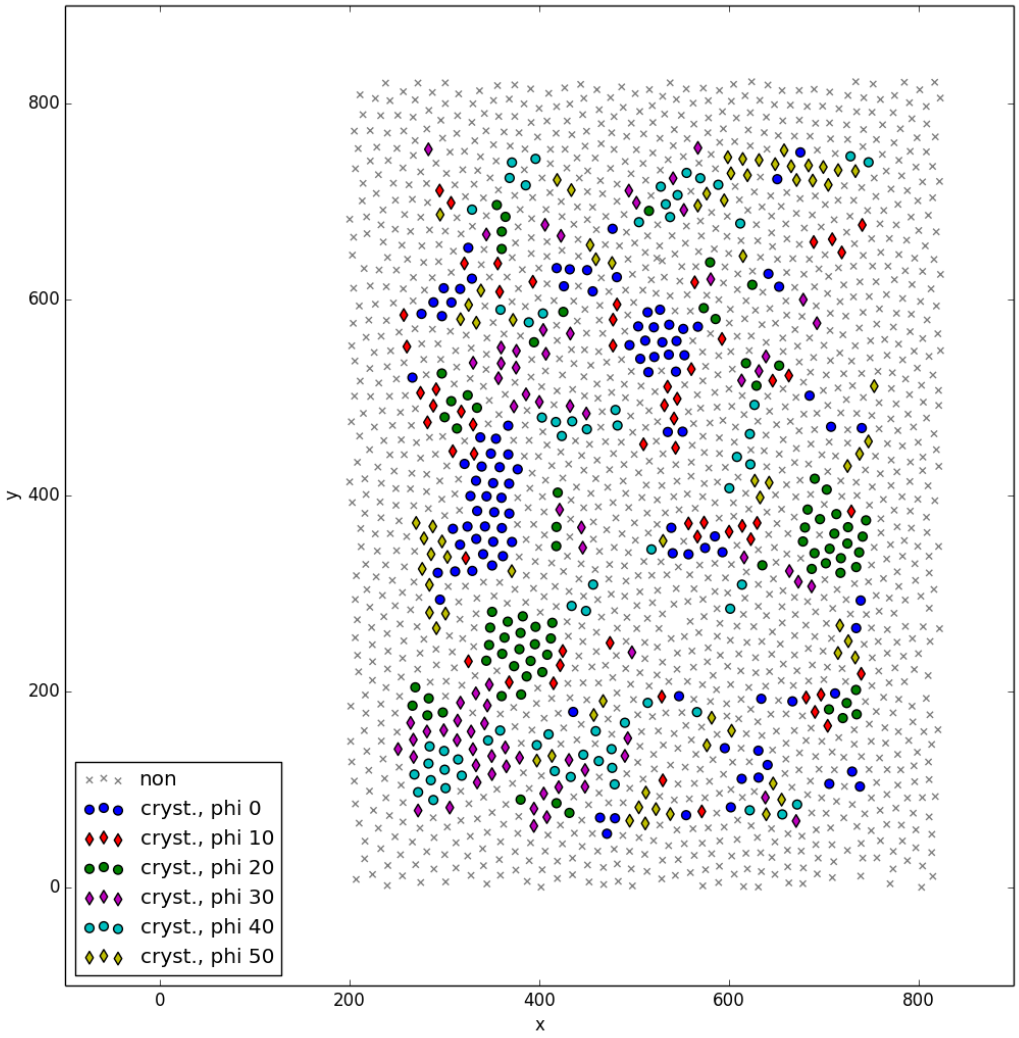

Figure 7.28: Recrystallization dataset, scan 720: After the electric pulse is switched off again, the particles are detectable again. They are very disordered at this point of time, but start to arrange slowly again.

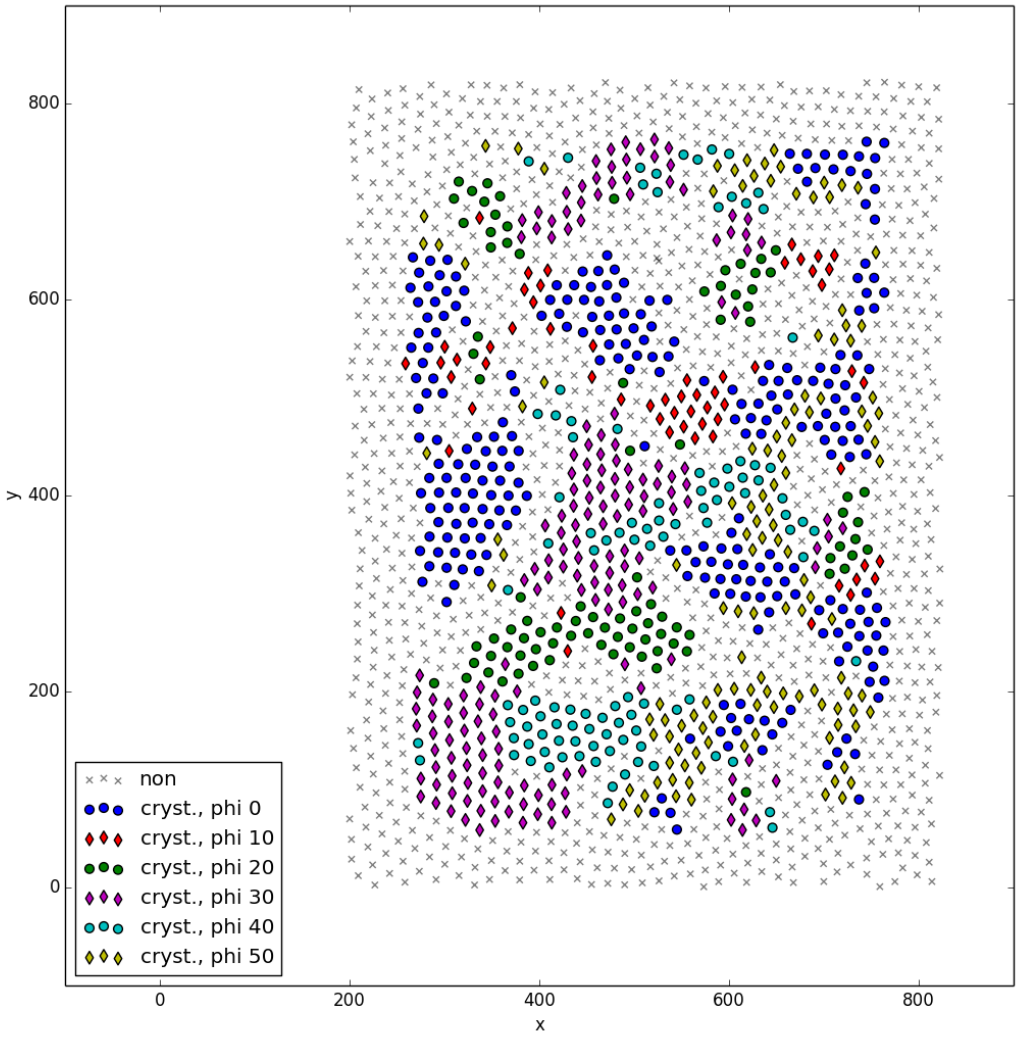

Figure 7.29: Recrystallization dataset, scan 800: the crystalline nuclei grow gradually, which have different orientations and are seperated by defect lines, but usually are quite homogeneous in their inner structure.

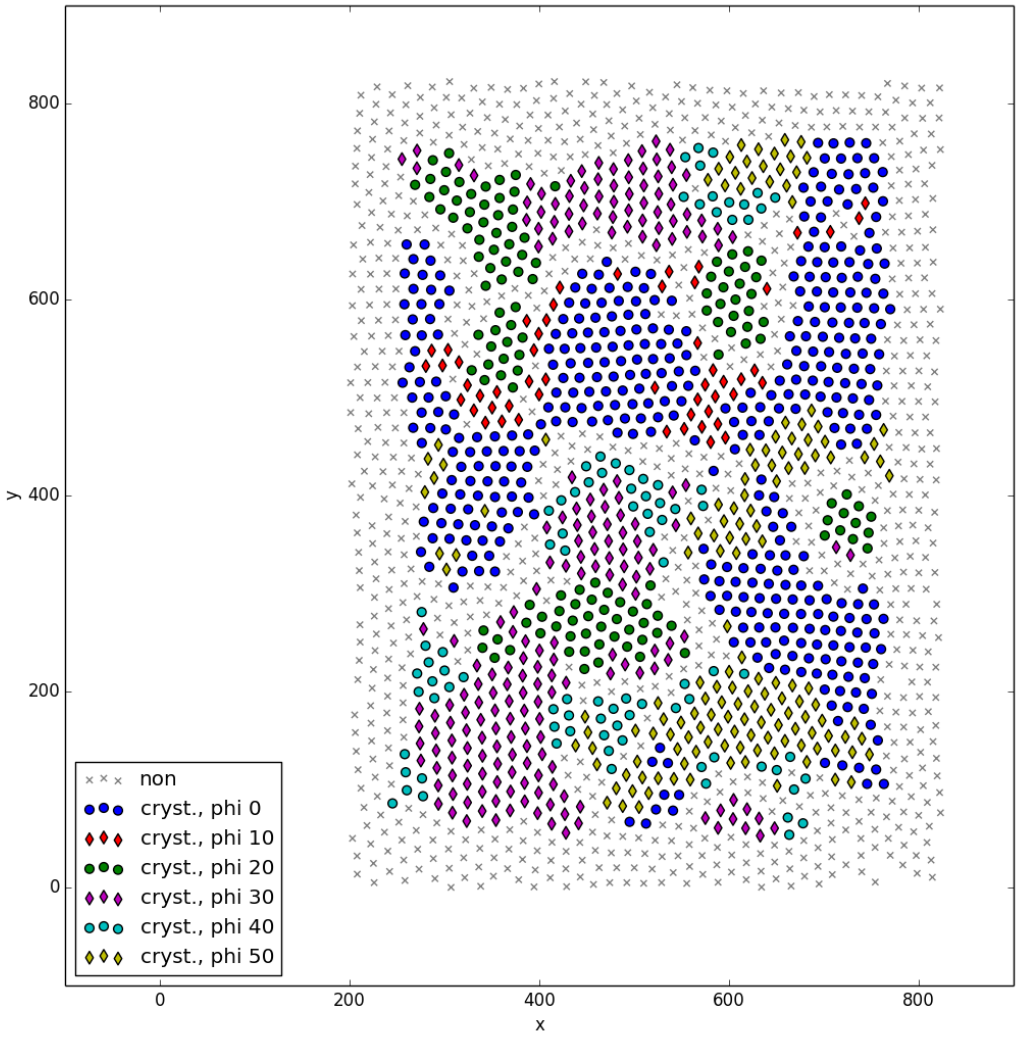

Figure 7.30: Recrystallization dataset, scan 900: the different areas grow together. The inner structure of the coalesced clusters are characterized by many orientational transitions of 10°.

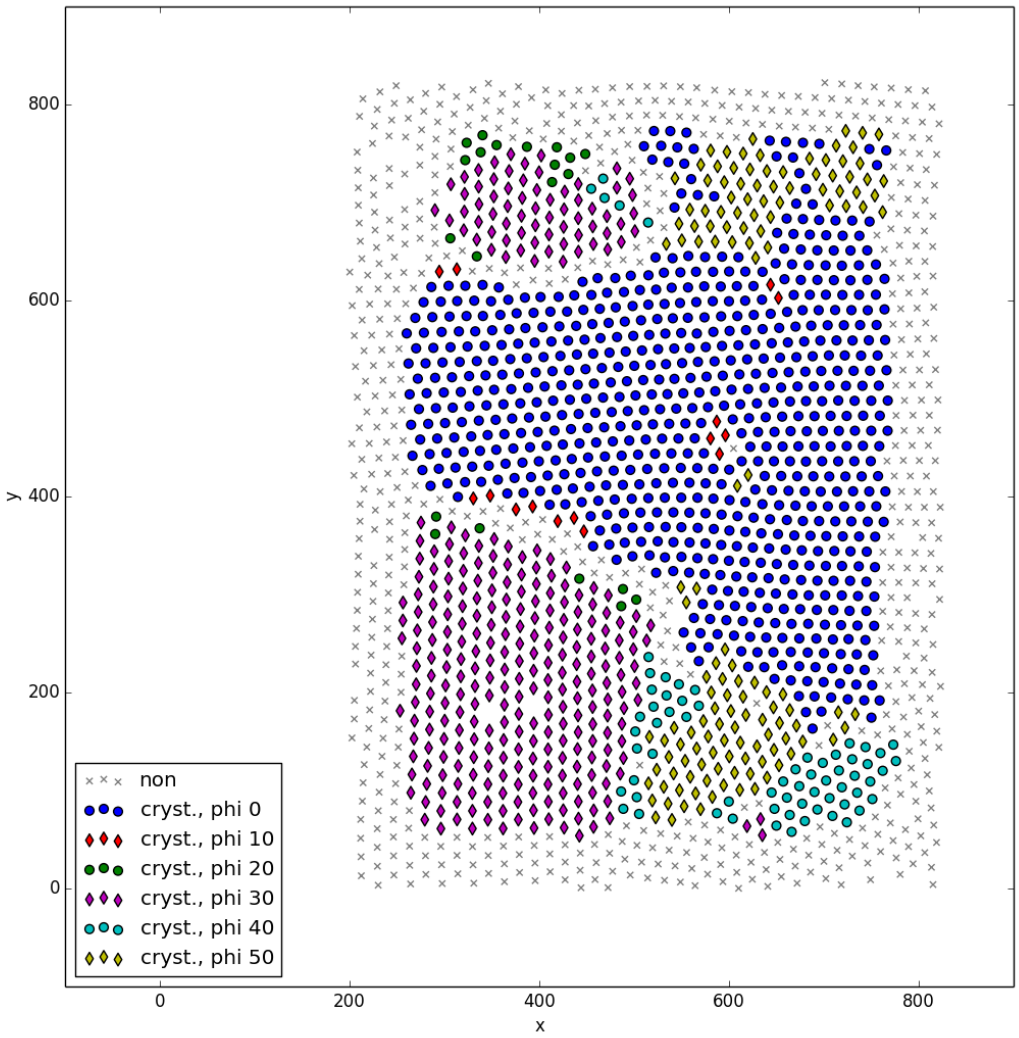

Figure 7.31: Recrystallization dataset, scan 2046: the last distribution in the set. The crystal has cooled down again and reached a quite stable shape. Its structure exists of two big and quite homogeneous clusters, seperated by a defect line in the middle, and two smaller and a bit more inhomogeneous clusters in the top left and the bottom right corner.

total number of particles N_{total}	16384
number of time steps N_{steps}	6000
time step dt	0.001383956027 s
initial particle distance	1.05 λ
frequency f_x	0.180 Hz
frequency f_y	0.190 Hz
frequency f_z	23.00 Hz
damping rate ν	1.48 $\frac{1}{s}$
temperature T	300 K
screening length λ	$4.00 \cdot 10^{-4}$ m
particle mass m	$6.10 \cdot 10^{-13}$ kg
particle charge Q	19100 e
ion wake "extra" charge q	0.20 Q
distance of "extra" charge δ	0.3 λ

Table 7.1: Parameters of the simulation by Ingo Laut [50] creating the "melting" dataset. The symbols for the parameters are directly comparable to [50].

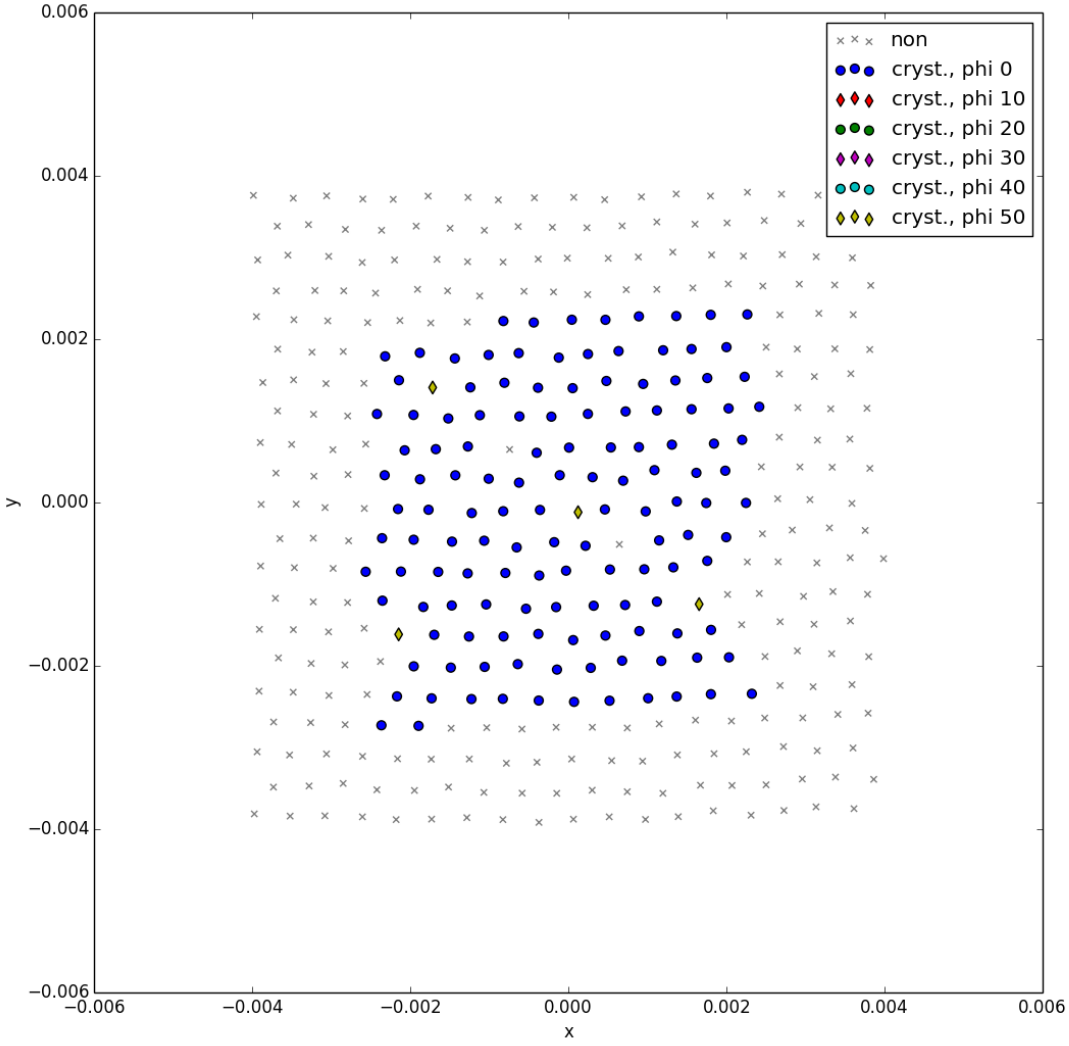

Figure 7.32: This plot shows the visualization of the results of scan 4700 of the dataset "melting". Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. Till scan 4700, although the level of motion has intensified, the shape of the crystal is still quite homogeneous. At his point of time, the crystal only is melting at the border region.

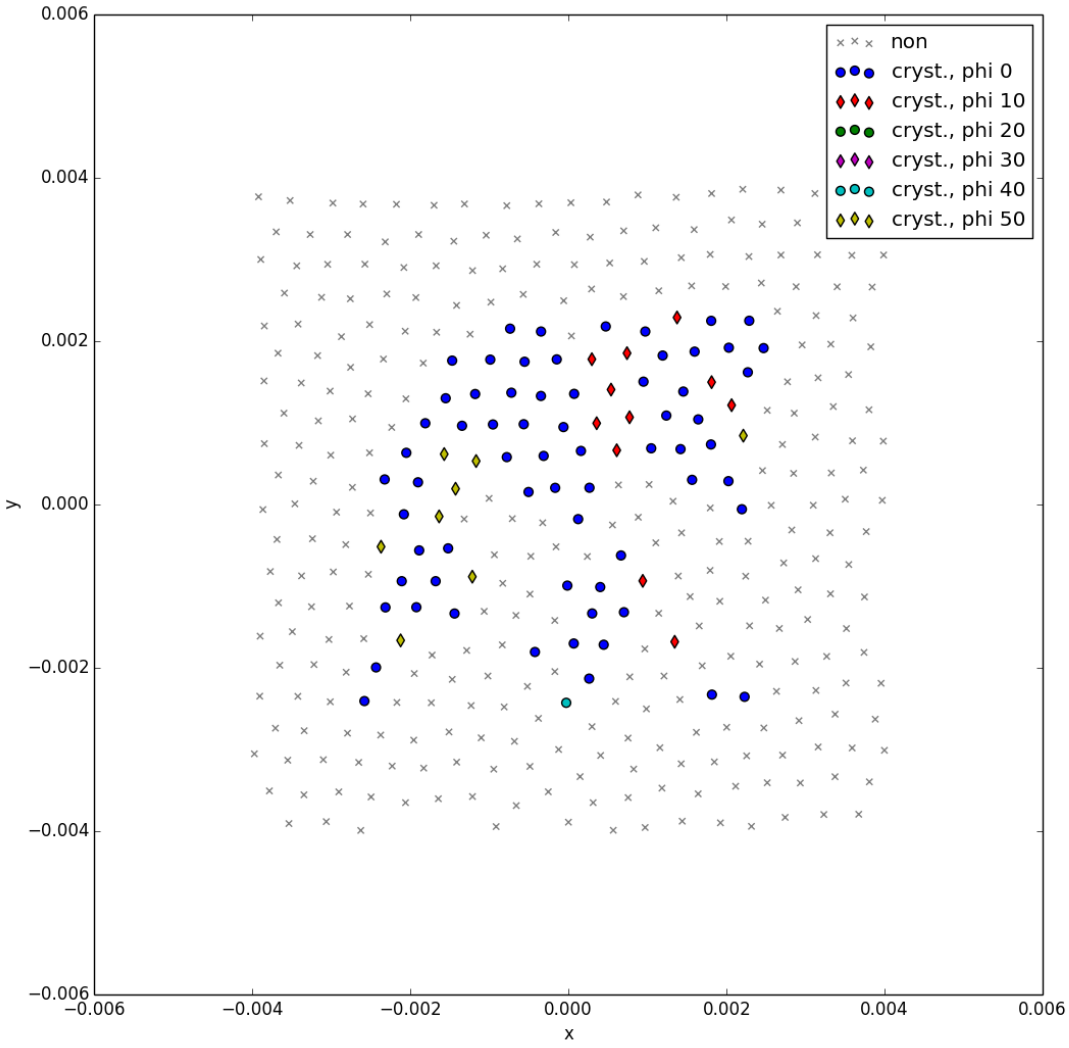

Figure 7.33: This plot shows the visualization of the results of scan 5000 of the dataset "melting". Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. In scan 5000, the level of motion has strongly advanced and gave rise to a partial melting of the crystal, also in the inner part.

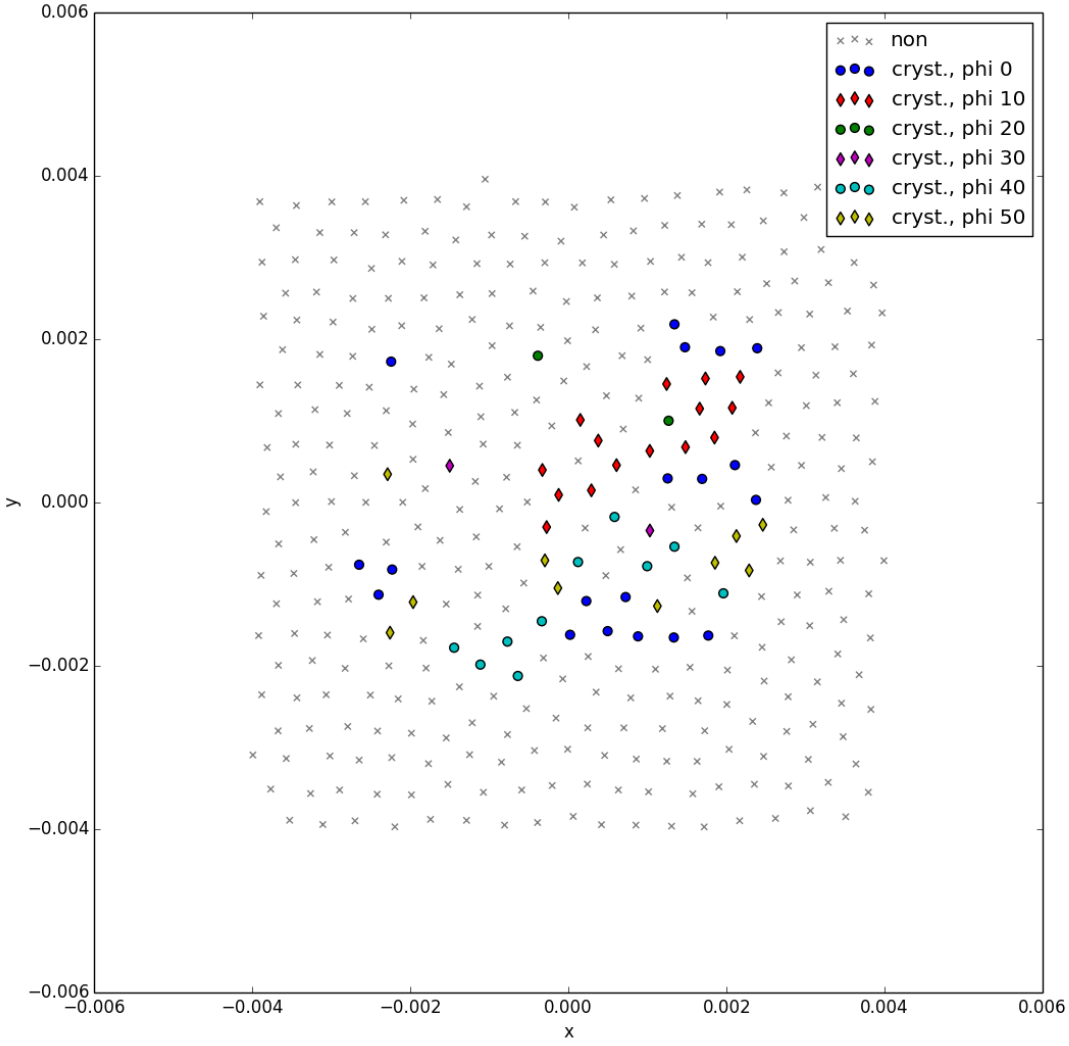

Figure 7.34: This plot shows the visualization of the results of scan 5400 of the dataset "melting". Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. Scan 5400 is one of the distributions with the smallest crystalline fraction. The fluctuations here are quite strong, the crystal is very inhomogeneous and all orientations are present.

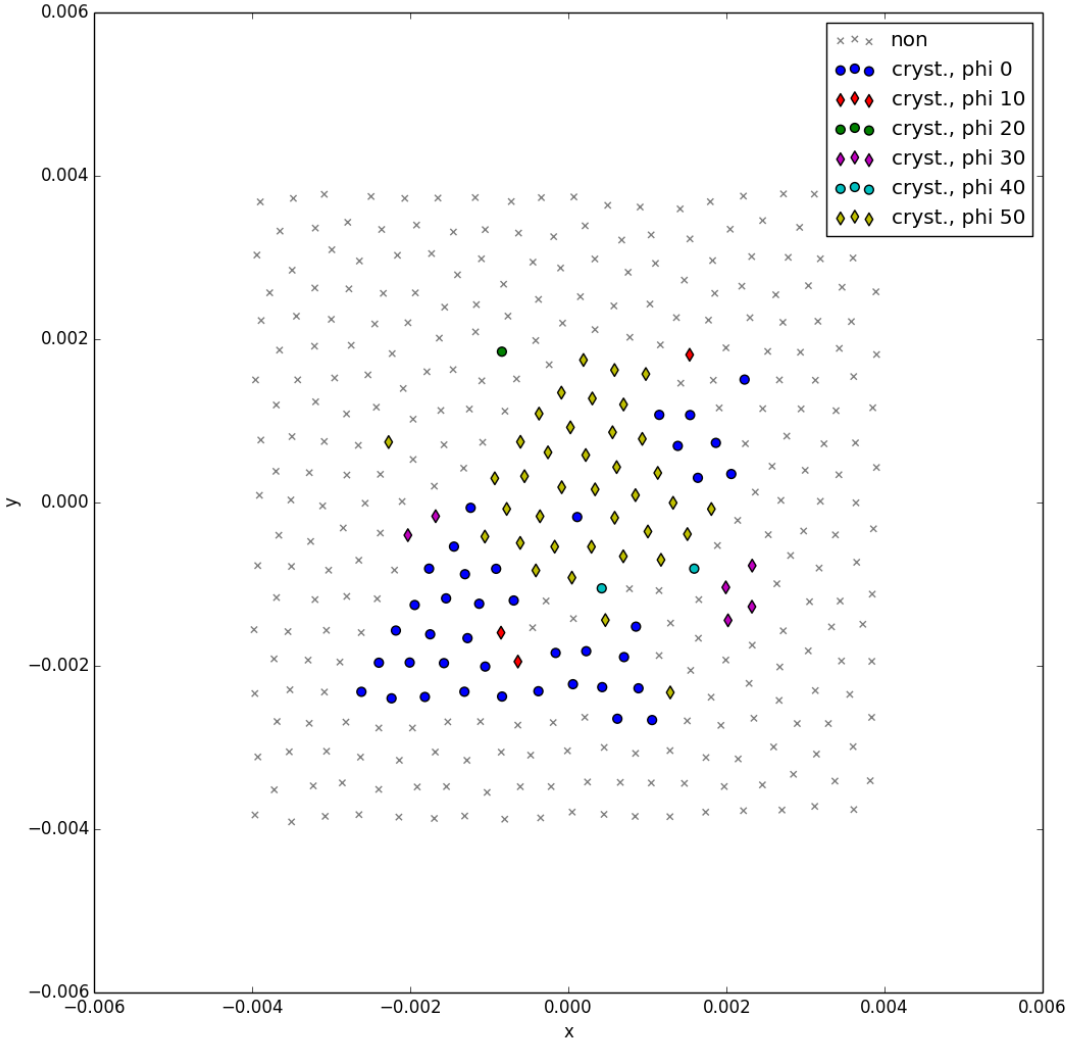

Figure 7.35: This plot shows the visualization of the results of scan 5999 of the dataset "melting". Different colors and shapes of the points refer to different orientations of the local structures, in the range $0°, 10° \ldots 50°$. Scan 5999 is the last one in the dataset. The size of the crystalline cluster has increased a bit again, due to the characteristics of the simulation [50], that creates oscillating alternations between melting and recrystallization.

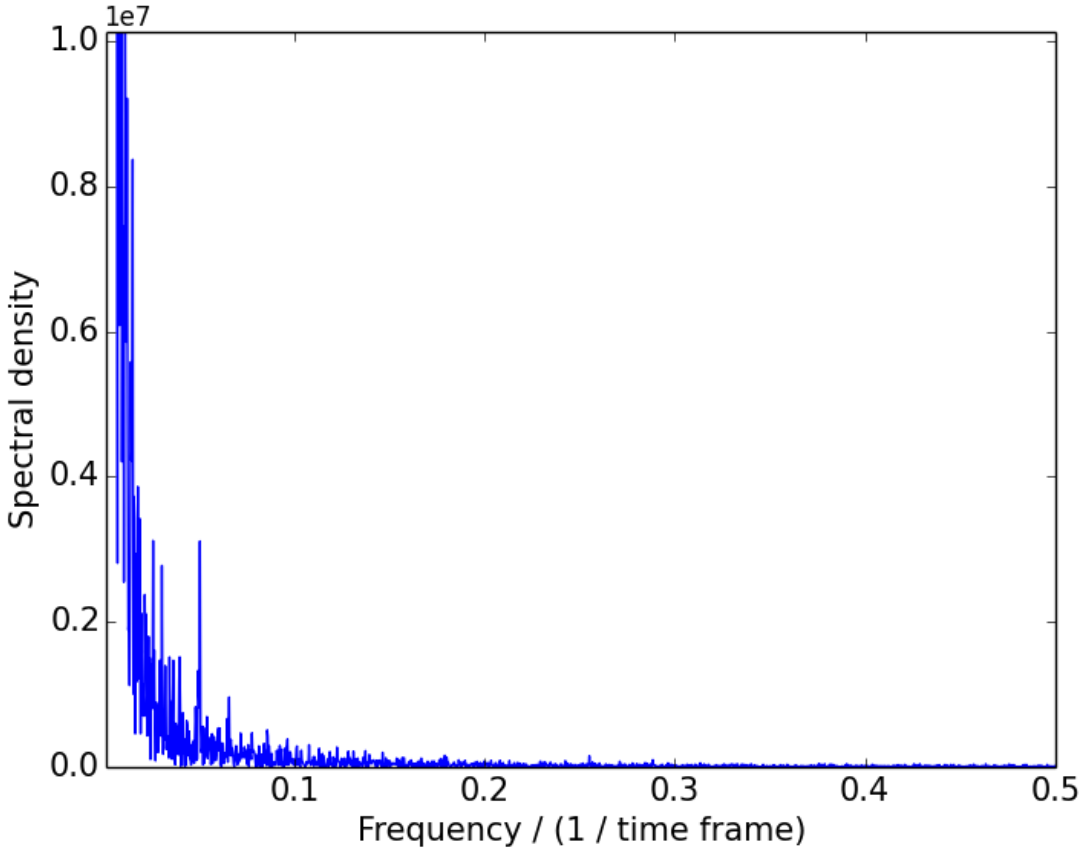

Figure 7.36: The power spectrum of the time series of the number of crystalline structures in the dataset "melting" with reference orientation 0 (blue graph in 6.3). The first, very large frequencies were cut out. The peak at frequency $0.05 \frac{1}{\text{time frame}}$ is easily observable in the graph in figure (6.3).

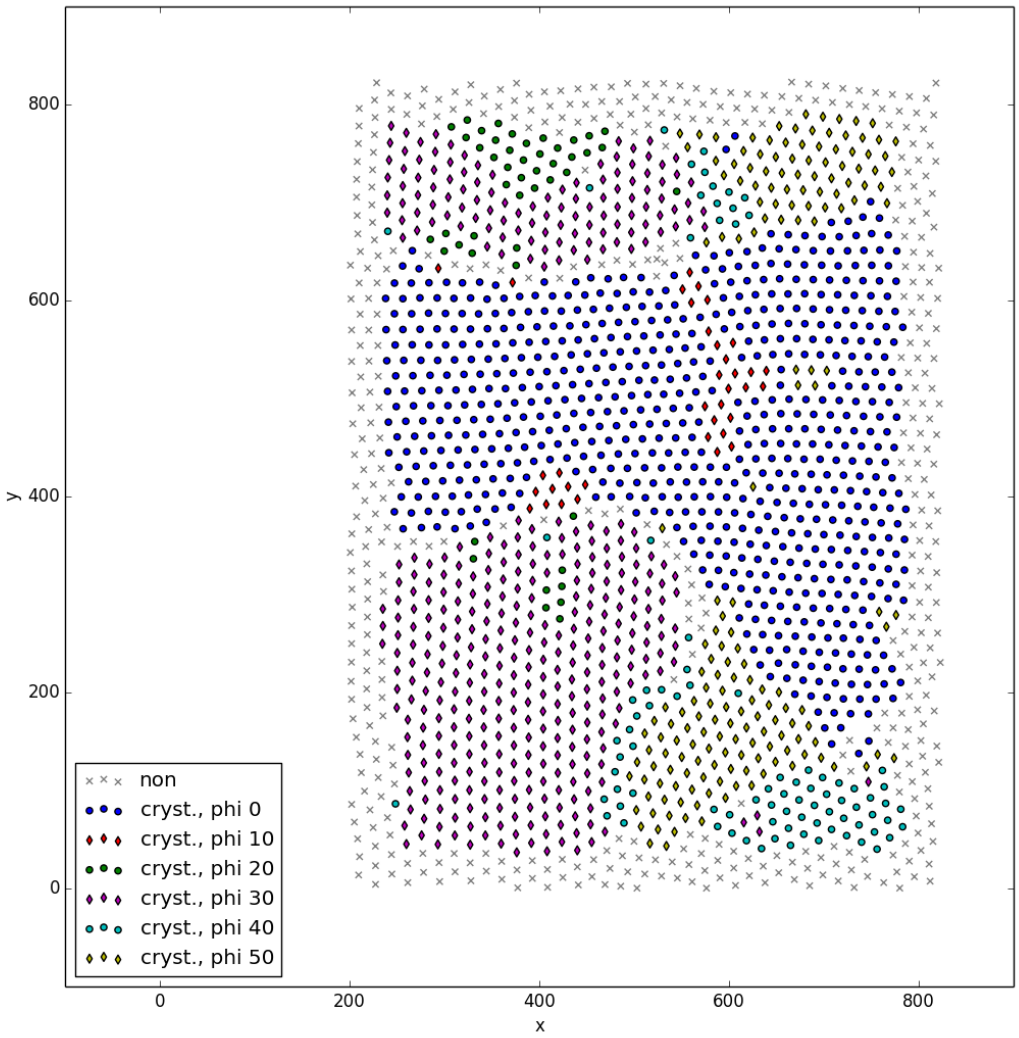

Figure 7.37: Demonstration of the edge detection of my algorithm by point distribution scan 1500 of the recrystallization dataset. The parameters used for this analysis are: $N_{neighbors} = 6$, normal *threshold*. With only $N_{neighbors} = 6$, the defect lines are barely observable.

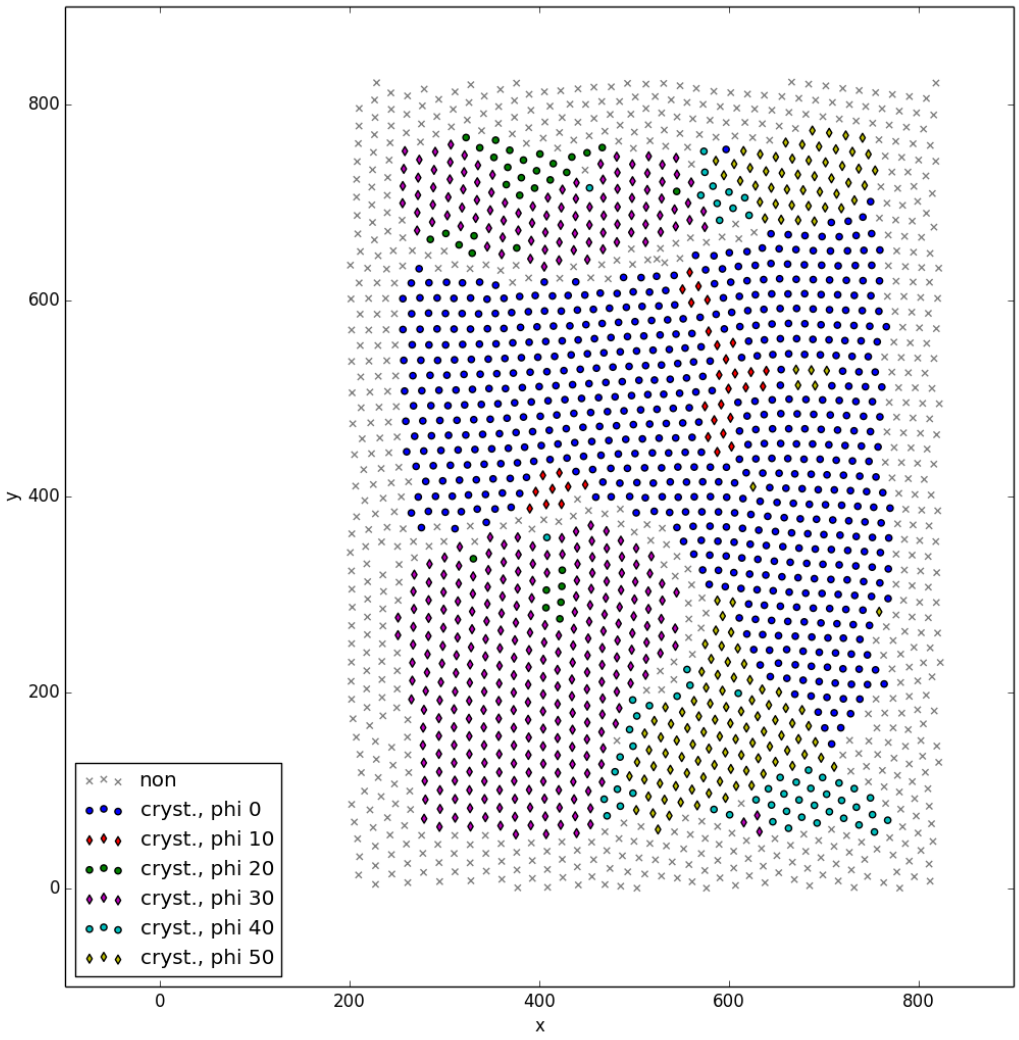

Figure 7.38: Demonstration of the edge detection of my algorithm by point distribution scan 1500 of the recrystallization dataset. The parameters used for this analysis are: $N_{neighbors} = 18$, normal *threshold*. With larger $N_{neighbors} = 18$, the defect lines get broadened. These value, $N_{neighbors} = 18$ and normal $threshold = 13.6$, are used in the algorithm for 2D point distributions.

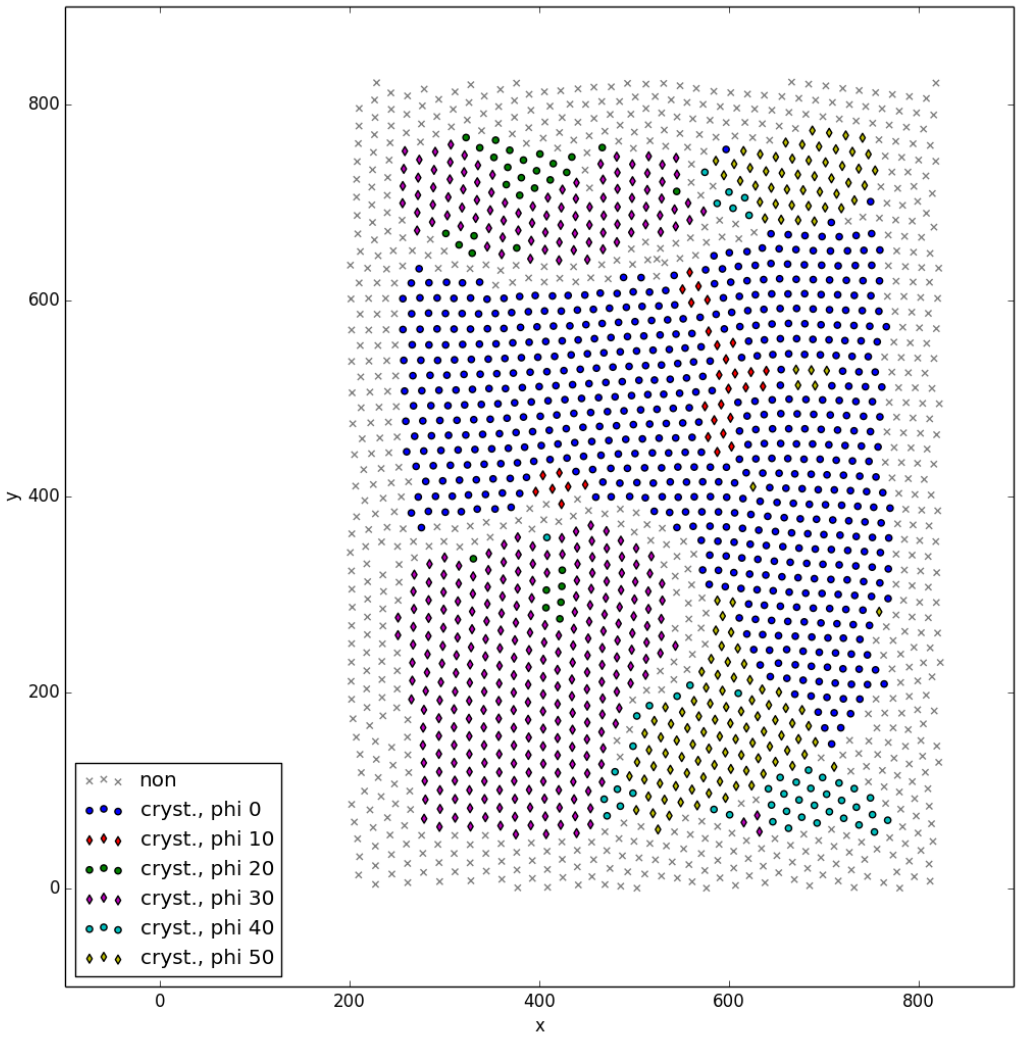

Figure 7.39: Demonstration of the edge detection of my algorithm by point distribution scan 1500 of the recrystallization dataset. The parameters used for this analysis are: $N_{neighbors} = 18$, lowered *threshold*. Lowering *threshold* has much more effect than increasing $N_{neighbors}$. The defect lines get much pronounced, especially with further lowered *threshold* (next figure).

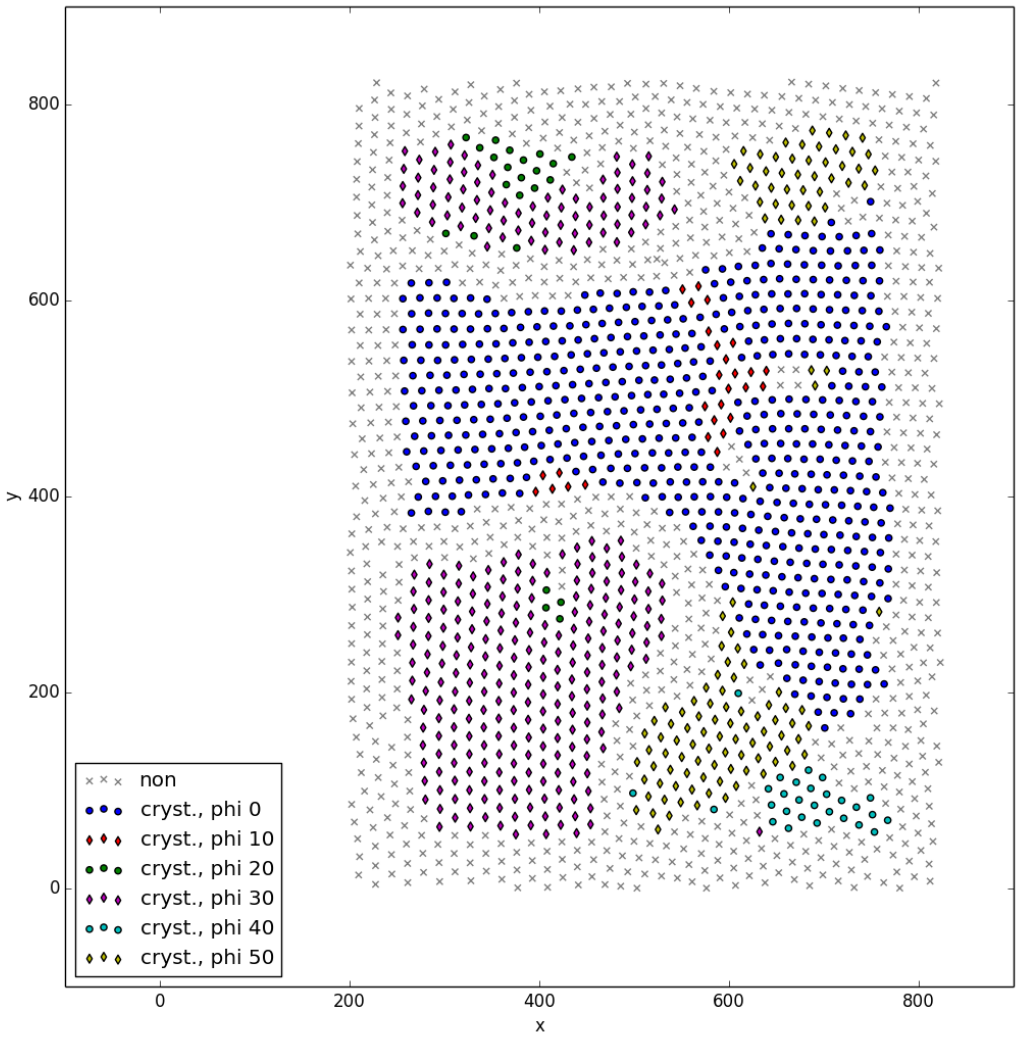

Figure 7.40: Demonstration of the edge detection of my algorithm by point distribution scan 1500 of the recrystallization dataset. The parameters used for this analysis are: $N_{neighbors} = 18$, further lowered *threshold*. The defect lines got even bolder, but the border region has not broadened.

Danksagung

Zuvorderst möchte ich Prof. Gregor Morfill und Dr. Christoph Räth danken, ihre Namen zur Verfügung gestellt und meine Betreuung übernommen zu haben. Besonders Christoph danke ich, immer sehr zugänglich zu sein und Kritik und Anregungen gleichzeitig angenehm und zweckmäßig zu präsentieren und zu diskutieren.

Weiterhin danke ich der Forschungsgruppe Komplexe Plasmen am DLR Oberpfaffenhofen, in der ich diese Arbeit anfertigte. Dr. Hubertus Thomas zu meiner Anstellung in der Gruppe, und vielen Personen für die anregenden Diskussionen und für die Bereitstellung ihrer Datensätze und Arbeiten. Besonders natürlich für die angenehme Atmosphäre und die vielen Gespräche, über Gott und die Welt, und auch mit vielen interessanten Einblicken in die Welt des Wissenschaftsbetriebs.

Christopher Dietz danke ich bereitwillig seine Daten und Ergebnisse zur Verfügung gestellt zu haben, auch für Präsentation in dieser Arbeit.

Zuletzt möchte ich die Studenten und Mitarbeiter dieser Arbeitsgruppe hervorheben, die nicht nur als Kollegen auftraten, sondern Teil des Privatlebens wurden.

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.